Anne Laurent
Olivier Strauss
Bernadette Bouchon-Meunier
Ronald R. Yager (Eds.)

Communications in Computer and Information Science 442

# Information Processing and Management of Uncertainty in Knowledge-Based Systems

15th International Conference, IPMU 2014
Montpellier, France, July 15-19, 2014
Proceedings, Part I

Part 1

Springer

# Communications in Computer and Information Science 442

Anne Laurent Olivier Strauss
Bernadette Bouchon-Meunier
Ronald R. Yager (Eds.)

# Information Processing and Management of Uncertainty in Knowledge-Based Systems

15th International Conference, IPMU 2014
Montpellier, France, July 15-19, 2014
Proceedings, Part I

Springer

Volume Editors

Anne Laurent
Université Montpellier 2, LIRMM, France
E-mail: laurent@lirmm.fr

Olivier Strauss
Université Montpellier 2, LIRMM, France
E-mail: olivier.strauss@lirmm.fr

Bernadette Bouchon-Meunier
Sorbonne Universités, UPMC Paris 6, France
E-mail: bernadette.bouchon-meunier@lip6.fr

Ronald R. Yager
Iona College, New Rochelle, NY, USA
E-mail: ryager@iona.edu

ISSN 1865-0929 e-ISSN 1865-0937
ISBN 978-3-319-08794-8 e-ISBN 978-3-319-08795-5
DOI 10.1007/978-3-319-08795-5
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014942459



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

Printed on acid-free paper

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

# Preface

Here we provide the proceedings of the 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, IPMU 2014, held in Montpellier, France, during July 15–19, 2014. The IPMU conference is organized every two years with the focus of bringing together scientists working on methods for the management of uncertainty and aggregation of information in intelligent systems.

This conference provides a medium for the exchange of ideas between theoreticians and practitioners working on the latest developments in these and other related areas. This was the 15th edition of the IPMU conference, which started in 1986 and has been held every two years in the following locations in Europe: Paris (1986), Urbino (1988), Paris (1990), Palma de Mallorca (1992), Paris (1994), Granada (1996), Paris (1998), Madrid (2000), Annecy (2002), Perugia (2004), Malaga (2008), Dortmund (2010) and Catania (2012).

Among the plenary speakers at past IPMU conferences, there have been three Nobel Prize winners: Kenneth Arrow, Daniel Kahneman, and Ilya Prigogine. An important feature of the IPMU Conference is the presentation of the Kampé de Fériet Award for outstanding contributions to the field of uncertainty. This year, the recipient was Vladimir N. Vapnik. Past winners of this prestigious award were Lotfi A. Zadeh (1992), Ilya Prigogine (1994), Toshiro Terano (1996), Kenneth Arrow (1998), Richard Jeffrey (2000), Arthur Dempster (2002), Janos Aczel (2004), Daniel Kahneman (2006), Enric Trillas (2008), James Bezdek (2010), Michio Sugeno (2012).

The program of the IPMU 2014 conference consisted of 5 invited academic talks together with 180 contributed papers, authored by researchers from 46 countries, including the regular track and 19 special sessions. The invited academic talks were given by the following distinguished researchers: Vladimir N. Vapnik (NEC Laboratories, USA), Stuart Russell (University of California, Berkeley, USA and University Pierre et Marie Curie, Paris, France), Inés Couso (University of Oviedo, Spain), Nadia Berthouze (University College London, United Kingdom) and Marcin Detyniecki (University Pierre and Marie Curie, Paris, France).

Industrial talks were given in complement of academic talks and highlighted the necessary collaboration we all have to foster in order to deal with current challenges from the real world such as Big Data for dealing with massive and complex data.

The success of IPMU 2014 was due to the hard work and dedication of a large number of people, and the collaboration of several institutions. We want to acknowledge the industrial sponsors, the help of the members of the International Program Committee, the reviewers of papers, the organizers of special sessions, the Local Organizing Committee, and the volunteer students. Most of all, we

appreciate the work and effort of those who contributed papers to the conference. All of them deserve many thanks for having helped to attain the goal of providing a high quality conference in a pleasant environment.

May 2014

Bernadette Bouchon-Meunier
Anne Laurent
Olivier Strauss
Ronald R. Yager

# Conference Committee

## General Chairs

| | |
|---|---|
| Anne Laurent | Université Montpellier 2, France |
| Olivier Strauss | Université Montpellier 2, France |

## Executive Directors

| | |
|---|---|
| Bernadette Bouchon-Meunier | CNRS-UPMC, France |
| Ronald R. Yager | Iona College, USA |

## Web Chair

| | |
|---|---|
| Yuan Lin | INRA, SupAgro, France |

## Proceedings Chair

| | |
|---|---|
| Jérôme Fortin | Université Montpellier 2, France |

## International Advisory Board

Giulianella Coletti (Italy)
Miguel Delgado (Spain)
Mario Fedrizzi (Italy)
Laurent Foulloy (France)
Salvatore Greco (Italy)
Julio Gutierrez-Rios (Spain)
Eyke Hüllermeier(Germany)
Luis Magdalena (Spain)
Christophe Marsala (France)
Benedetto Matarazzo (Italy)
Manuel Ojeda-Aciego (Spain)
Maria Rifqi (France)
Lorenza Saitta (Italy)
Enric Trillas (Spain)
Llorenç Valverde (Spain)
José Luis Verdegay (Spain)
Maria-Amparo Vila (Spain)
Lotfi A. Zadeh (USA)

## Special Session Organizers

| | |
|---|---|
| Jose M. Alonso | European Centre for Soft Computing, Spain |
| Michal Baczynski | University of Silesia, Poland |
| Edurne Barrenechea | Universidad Pública de Navarra, Spain |
| Zohra Bellahsene | University Montpellier 2, France |
| Patrice Buche | INRA, France |
| Thomas Burger | CEA, France |
| Tomasa Calvo | Universidad de Alcalá, Spain |

| | |
|---|---|
| Brigitte Charnomordic | INRA, France |
| Didier Coquin | University of Savoie, France |
| Cécile Coulon-Leroy | École Supérieure d'Agriculture d'Angers, France |
| Sébastien Destercke | CNRS, Heudiasyc Lab., France |
| Susana Irene Diaz Rodriguez | University of Oviedo, Spain |
| Luka Eciolaza | European Centre for Soft Computing, Spain |
| Francesc Esteva | IIIA - CSIC, Spain |
| Tommaso Flaminio | University of Insubria, Italy |
| Brunella Gerla | University of Insubria, Italy |
| Manuel González Hidalgo | University of the Balearic Islands, Spain |
| Michel Grabisch | University of Paris Sorbonne, France |
| Serge Guillaume | Irstea, France |
| Anne Laurent | University Montpellier 2, France |
| Christophe Labreuche | Thales Research and Technology, France |
| Kevin Loquin | University Montpellier 2, France |
| Nicolás Marín | University of Granada, Spain |
| Trevor Martin | University of Bristol, UK |
| Sebastia Massanet | University of the Balearic Islands, Spain |
| Juan Miguel Medina | University of Granada, Spain |
| Enrique Miranda | University of Oviedo, Spain |
| Javier Montero | Complutense University of Madrid, Spain |
| Jesús Medina Moreno | University of Cádiz, Spain |
| Manuel Ojeda Aciego | University of Málaga, Spain |
| Martin Pereira Farina | Universidad de Santiago de Compostela, Spain |
| Olga Pons | University of Granada, Spain |
| Dragan Radojevic | Serbia |
| Anca L. Ralescu | University of Cincinnati, USA |
| François Scharffe | University Montpellier 2, France |
| Rudolf Seising | European Centre for Soft Computing, Spain |
| Marco Elio Tabacchi | Università degli Studi di Palermo, Italy |
| Tadanari Taniguchi | Tokai University, Japan |
| Charles Tijus | LUTIN-CHArt, Université Paris 8, France |
| Konstantin Todorov | University Montpellier 2, France |
| Lionel Valet | University of Savoie, France |

## Program Committee

| | | |
|---|---|---|
| Michal Baczynski | Ulrich Bodenhofer | Rita Casadio |
| Gleb Beliakov | P. Bonissone | Yurilev Chalco-Cano |
| Radim Belohlavek | Bernadette Bouchon-Meunier | Brigitte Charnomordic |
| Salem Benferhat | Patrice Buche | Guoqing Chen |
| H. Berenji | Humberto Bustince | Carlos A. Coello Coello |
| Isabelle Bloch | | Giulianella Coletti |

Oscar Cordón
Inés Couso
Keeley Crockett
Bernard De Baets
Sébastien Destercke
Marcin Detyniecki
Antonio Di Nola
Gerard Dray
Dimiter Driankov
Didier Dubois
Francesc Esteva
Mario Fedrizzi
Janos Fodor
David Fogel
Jérôme Fortin
Laurent Foulloy
Sylvie Galichet
Patrick Gallinari
Maria Angeles Gil
Siegfried Gottwald
Salvatore Greco
Steve Grossberg
Serge Guillaume
Lawrence Hall
Francisco Herrera
Enrique Herrera-Viedma
Kaoru Hirota
Janusz Kacprzyk
A. Kandel
James M. Keller
E.P. Klement
Laszlo T. Koczy
Vladik Kreinovich
Christophe Labreuche
Jérôme Lang
Henrik L. Larsen
Anne Laurent
Marie-Jeanne Lesot
Churn-Jung Liau
Yuan Lin
Honghai Liu
Weldon Lodwick
Kevin Loquin
Thomas Lukasiewicz
Luis Magdalena
Jean-Luc Marichal
Nicolás Marín
Christophe Marsala
Trevor Martin
Mylène Masson
Silvia Massruha
Gilles Mauris
Gaspar Mayor
Juan Miguel Medina
Jerry Mendel
Enrique Miranda
Pedro Miranda
Paul-André Monney
Franco Montagna
Javier Montero
Jacky Montmain
Serafín Moral
Yusuke Nojima
Vilem Novak
Hannu Nurmi
S. Ovchinnikov
Nikhil R. Pal
Endre Pap
Simon Parsons
Gabriella Pasi
W. Pedrycz
Fred Petry
Vincenzo Piuri
Olivier Pivert
Henri Prade
Anca Ralescu
Dan Ralescu
Mohammed Ramdani
Marek Reformat
Agnes Rico
Maria Rifqi
Enrique H. Ruspini
Glen Shafer
J. Shawe-Taylor
P. Shenoy
P. Sobrevilla
Umberto Straccia
Olivier Strauss
M. Sugeno
Kay-Chen Tan
S. Termini
Konstantin Todorov
Vicenc Torra
I. Burhan Turksen
Barbara Vantaggi
José Luis Verdegay
Thomas Vetterlein
Marco Zaffalon
Hans-Jürgen Zimmermann
Jacek Zurada

## Additional Members of the Reviewing Committee

Iris Ada
Jesús Alcalá-Fernandez
Cristina Alcalde
José María Alonso
Alberto Alvarez-Alvarez
Leila Amgoud
Massih-Reza Amini
Derek Anderson
Plamen Angelov
Violaine Antoine
Alessandro Antonucci
Alain Appriou
Arnaud Martin
Jamal Atif
Rosangela Ballini
Carlos Barranco
Edurne Barrenechea
Gleb Beliakov
Zohra Bellahsene
Alessio Benavoli
Eric Benoit

Moulin Bernard
Christophe Billiet
Fernando Bobillo
Gloria Bordogna
Christian Borgelt
Stefan Borgwardt
Carole Bouchard
Antoon Bronselaer
Thomas Burger
Ana Burusco
Manuela Busaniche
Silvia Calegari
Tomasa Calvo
Jesus Campaña
Andrés Cano
Philippe Capet
Arnaud Castelltort
Laëtitia Chapel
Yi-Ting Chiang
Francisco Chiclana
Laurence Cholvy
Davide Ciucci
Chloé Clavel
Ana Colubi
Arianna Consiglio
Didier Coquin
Pablo Cordero
María Eugenia Cornejo
Chris Cornelis
Cécile Coulon-Leroy
Fabio Cozman
Madalina Croitoru
Jc Cubero
Fabio Cuzzolin
Inmaculada de Las Peñas Cabrera
Gert Decooman
Afef Denguir
Thierry Denœux
Lynn D'Eer
Guy De Tré
Irene Diaz Rodriguez
Juliette Dibie-Barthélémy
Pawel Drygaś
Jozo Dujmovic
Fabrizio Durante
Luka Eciolaza
Nicola Fanizzi
Christian Fermüller
Javier Fernandez
Tommaso Flaminio
Carl Frélicot
Ramón Fuentes-González
Louis Gacogne
Jose Galindo
José Luis García-Lapresta
Brunella Gerla
Maria Angeles Gil
Lluis Godo
Manuel Gonzalez
Michel Grabisch
Salvatore Greco
Przemyslaw Grzegorzewski
Kevin Guelton
Thierry Marie Guerra
Robert Hable
Allel Hadjali
Pascal Held
Sascha Henzgen
Hykel Hosni
Celine Hudelot
Julien Hué
Eyke Hüllermeier
Atsushi Inoue
Uzay Kaymak
Anna Kolesarova
Jan Konecny
Stanislav Krajci
Ondrej Kridlo
Rudolf Kruse
Naoyuki Kubota
Maria Teresa Lamata
Cosmin Lazar
Florence Le Ber
Fabien Lehuede
Ludovic Lietard
Berrahou Lilia
Chin-Teng Lin
Xinwang Liu
Nicolas Madrid
Ricardo A. Marques Pereira
Luis Martinez
Carmen Martinez-Cruz
Sebastià Massanet
Brice Mayag
Gaspar Mayor
Jesús Medina Moreno
Corrado Mencar
David Mercier
Radko Mesiar
Andrea Mesiarova
Arnau Mir
Franco Montagna
Susana Montes
Tommaso Moraschini
María Moreno García
Olivier Naud
Manuel Ojeda-Aciego
Daniel Paternain
Andrea Pedrini
Carmen Peláez
Martín Pereira-Farina
David Perez
Irina Perfilieva
Nathalie Perrot
Valerio Perticone
Davide Petturiti
Pascal Poncelet
Olga Pons
Thierry Pun
Erik Quaeghebeur
Benjamin Quost
Dragan Radojevic
Eloisa Ramírez
Jordi Recasens
Juan Vicente Riera
Antoine Rolland
Daniel Ruiz
M. Dolores Ruiz
Nobusumi Sagara

Daniel Sanchez
Sandra Sandri
Francois Scharffe
Johan Schubert
Rudolf Seising
Carlo Sempi
Robin Senge
Ammar Shaker
Prakash Shenoy
Beatriz Sinova
Damjan Skulj
Dominik Slezak
Roman Slowinski
Alejandro Sobrino
Hélène Soubaras
Jaume Suner
Nicolas Sutton-Charani
Ronan Symoneaux
Eulalia Szmidt
Marcoelio Tabacchi
Eiichiro Takahagi
Nouredine Tamani
Tadanari Taniguchi
Elizabeth Tapia
Joao Manuel R.S. Tavares
Maguelonne Teisseire
Matthias Thimm
Rallou Thomopoulos
Charles Tijus
Joan Torrens
Carmen Torres
Gwladys Toulemonde
Nicolas Tremblay
Gracián Trivino
Matthias Troffaes
Luigi Troiano
Wolfgang Trutschnig
Esko Turunen
Lionel Valet
Francisco Valverde
Arthur Vancamp
Barbara Vantaggi
Jirina Vejnarova
Paolo Vicig
Amparo Vila
Pablo Villacorta
Serena Villata
Sofía Vis
Paul Weng
Dongrui Wu
Slawomir Zadrozny

# Invited Talks

# Learning with Nontrivial Teacher

Vladimir Vapnik

NEC Laboratories, Princeton, New Jersey
Columbia University, New York City
U.S. National Academy of Engineering

**Abstract.** In this talk I will discuss a new learning paradigm which uses an intelligent agent (nontrivial teacher), the so called Learning Using Privileged Information (LUPI) paradigm. In the LUPI learning paradigm in order to machine find a good decision rule using small number of training examples (say in pattern recognition problem) teacher during training session adds to the training examples some additional information (privileged information) such as comments, explanations, metaphoric reasoning, and so on. This information will not be available for testing. I will discuss the property of new learning model and the role of nontrivial teacher it relation to the general problem of inference and construction intelligent (that use different from brute force methods) machines.

# New Challenges in Fuzzy Reasoning: When Practice and Theory Meet

Marcin Detyniecki

CNRS, UMR 7606, LIP6, Paris, France
Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, Paris, France
Polish Academy of Sciences, IBS PAN, Warsaw, Poland

**Abstract.** Fuzzy reasoning is a seminal tool for intelligent systems. It enables handling imprecise information and uncertain knowledge in real-world applications. Fuzzy reasoning can be seen as a process that finds a solution for a system of relational assignment equations. Generally, it is formulated as a compositional rule of inference, based on the modus ponens as deduction mechanism. This type of approximate reasoning is typically exploited in fuzzy rule-based systems. This general framework has played a key role in the success of fuzzy logic in general, and in particular, in the eyes of the general public with a number of intelligent applications.

Although approximate inference mechanisms and the underlying logic have been intensively studied, we observe that theoretical results do not necessarily address the challenges found in applications. Neither have the applied researchers integrated formal results in the engineering solution. This contribution aims at bringing some light to this matter.

In 1994, in a plenary talk titled "Fuzzy Logic and Soft Computing: Issues, Contentions and Perspectives", Lotfi Zadeh attributed the criticism towards fuzzy logic to a misunderstanding of what it is. He argued that fuzzy logic is used in two different senses: the narrow one, in which it is considered as a logical system which is an extension of multivalued logic; the broader sense, in which it is considered from a set-theory point of view as theory of classes with unsharp boundaries. At that time Zadeh pointed out that the success in terms of application, measured by the rapid growth in their number and their variety, is due to the fact that fuzzy logic in the broad sense has not only an enhanced ability to model real world problems thanks to its generality and expressive power, but also a capacity to achieve tractability and robustness with low solution cost. He also pointed out that, although there has been a certain progress in the understanding of this logic, the foundations needed to become firmer and its impact within mathematics more substantive.

Twenty years later, we observe that the theoretical foundations of fuzzy logic have been extremely well addressed, in particular in the logical strict sense. In parallel, fuzzy reasoning in the broad sense has been able to address complex issues in control, data analysis, and natural knowledge management in several application domains. Today, in order to take

fuzzy logic, and in particular fuzzy reasoning, to next level of performance, we need a better mutual understanding and interaction between researchers focusing on abstract theoretical structures and those focusing on operational intelligent systems. Twenty years ago Zadeh claimed that we need more foundations for fuzzy logic in the narrow sense. I advocate that now is time to reinforce the foundations of the *broad* fuzzy logic.

In this presentation centred on fuzzy reasoning, I will illustrate the gap that exists between available theoretical explanations, which are *de facto* fuzzy logic in the narrow sense, and practitioners expectations, who exploit fuzzy logic in a broad sense. In particular, I will use some successful applications in *multimedia information management* to show some paradoxical behaviours that may occur, *in practice*, when using standard fuzzy inference mechanisms. I will analyse these exemples from different perspectives by, in the one hand, formalising the underlying practical abstract motivations, and in the other hand, by providing theoretical explanations of the existing limitations. Moreover, I will show preliminary works - neither purely practical, neither theoretical - that provide some answers to the observed behaviours. These examples are intended to open some new tracks between the narrow and the broad fuzzy logic, around which the applied and the theoretical researcher can put up to start an exchange and a mutual understanding.

**Acknowledgments.** Results and reflections presented here are the result of a long collective work with different researchers from around the world, each having its own theoretical or applicative perspective. Without forgetting all with whom I have had enlightening discussions, I would to like to acknowledge my gratitude in particular to Sabrina Tollari, Marie-Jeanne Lesot, Maria Rifqi, Bernadette Bouchon-Meunier, Nicolas Labroche, Adrien Revault d'Allonnes, Christophe Marsala and all, my present and past, doctorate students.

# What Does Your Body Tell Me...., Oh, ... and What Does It Tell You?

Nadia Berthouze

UCLIC, University College London, London, UK
n.berthouze@ucl.ac.uk

**Abstract.** Recent years have seen the emergence of technology that involves and requires users to be engaged through their body. This has opened the possibility to better understand and exploit this modality to capture, respond to and regulate users' affective experience. Indeed, various studies in psychology have shown that our posture and body movement do communicate to others how we feel, and the intensity with which we feel, to the same extent that facial expressions and tone of voice do (for a review, see [1]). At the same time, studies have shown that body expressions are not only a window into one's emotional experience but they are also a mean to affect and regulate this experience (for a review, see [2]): the way we stand and the way we move affect our emotional state, our cognitive abilities and our attitude towards the events and environment that surround us and that we evaluate. In my talk, I will present a set of case studies that aim to show how these two sides of bodily expression can be used to design better and more effective technology.

Using applications in computer games [3,4], in physical rehabilitation [5] and in human-avatar interaction [6], I will show how body expressions can be used to automatically capture how the person feels. In doing so, I will also discuss the challenges raised by collecting, labelling and modelling naturalistic body expressions, as well as possible directions to address them. I will also bring into light how the availability of new devices and sensors make it possible to extend the study of body to aspects of it that are still under-investigated: touch behaviour as an extension of body behaviour [7] and muscle activity as the engine of body expressions.

Using the above case studies, I will discuss how this information could be used to adapt technology to better respond to the need of the user or of the goal to be achieved [8]. In particular, I will discuss how tracking the emotional states of the person through their body expressions offers new possibilities for self-directed rehabilitation. Emotional states are often forgotten when rehabilitation technology is designed with the main focus being on correcting movement. Unfortunately, correcting movement is not all what physical rehabilitation is about [9].

Finally, I will conclude by discussing how designers of full body technology could take advantage of what the body has to offer. I will present psychology and HCI research that aims to provide a principled approach

to the design of body movement to steer the user's emotional states towards what is most appropriate to the accomplishment of the task at hand [2,10].

## References


1. Kleinsmith, A., Bianchi-Berthouze, N.: Affective Body Expression Perception and Recognition: A Survey. IEEE Transactions on Affective Computing 4(1), 15–33 (2012)
2. Bianchi-Berthouze, N.: Understanding the role of body movement in player engagement. Human-Computer Interaction 28, 1–36 (2013)
3. Kleinsmith, A., Bianchi-Berthouze, N., Steed, A.: Automatic Recognition of Non-Acted Affective Postures. IEEE Trans Systems, Man and Cybernetics, Part B: Cybernteics 41(4), 1027–1038 (2011)
4. Savva, N., Scarinzi, A., Bianchi-Berthouze, N.: Continuous Recognition of Player's Affective Body Expression as Dynamic Quality of Aesthetic Experience. IEEE Trans. on Comp. Intell. AI in Games. 4(3), 199–212 (2012)
5. Aung, M.S.H., Bianchi-Berthouze, N., Williams, A., CdeC, W.P.: Automatic Recognition of Fear-Avoidance Behahviour in Chronic Pain Physical Rehabiliation. In: Proc. Pervasive Health (2014)
6. Griffin, H.J., Aung, M.S.H., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., Bianchi-Berthouze, N.: Laughter type recognition from whole body motion. In: Proc. Affective Computing and Intelligent Interaction, pp. 349–355 (2013)
7. Gao, Y., Bianchi-Berthouze, N., Meng, H.: What does touch tell us about emotions in touchscreen-based gameplay? TOCHI 31 (2012)
8. Nijhar, J., Bianchi-berthouze, N., Boguslawski, G.: Does Movement Recognition Precision affect the Player Experience in Exertion Games? In: International Conference on Intelligent Technologies for Interactive Entertainment. LNICST, vol. 78, pp. 73–82 (2012)
9. Singh, A., Klapper, A., Jia, J., Fidalgo, A., TajaduraJimenez, A., Kanakam, N., Williams, A., Bianchi-Berthouze, N.: Opportunities for technology in motivating people with chronic pain to do physical activity. In: CHI 2014 (2014)
10. Pasch, M., Bianchi-Berthouze, N., van Dijk, B., Nijholt, A.: Movement-based Sports Video Games: Investigating Motivation and Gaming Experience. Entertainment Computing 9(2), 169–180 (2009)

# Table of Contents – Part I

## Fuzzy Implications



## Clustering

## Fuzzy Measures and Integrals



## Non-Classical Logics

## Data Analysis



## Real-World Applications

## Aggregation



## Probabilistic Networks



## Recommendation Systems and Social Networks

## Fuzzy Systems

# Table of Contents – Part II

## Fuzzy Logic in Boolean Framework



## Management of Uncertainty in Social Networks



## From Different to Same, from Imitation to Analogy

## Soft Computing and Sensory Analysis



## Database Systems

## Fuzzy Set Theory



## Measurement and Sensory Information



## Aggregation

## Formal Methods for Vagueness and Uncertainty in a Many-Valued Realm



## Graduality



## Preferences

## Uncertainty Management in Machine Learning



## Philosophy and History of Soft Computing

## Soft Computing and Sensory Analysis

# Table of Contents – Part III

## Similarity Analysis



## Fuzzy Logic, Formal Concept Analysis and Rough Set

## Intelligent Databases and Information Systems



## Theory of Evidence

## Aggregation Functions

## Big Data - The Role of Fuzzy Methods



## Imprecise Probabilities: From Foundations to Applications

## Intelligent Measurement and Control for Nonlinear Systems



## Aggregation Functions

# Preference Relations and Families of Probabilities: Different Sides of the Same Coin

Inés Couso*

Department of Statistics and O.R.
University of Oviedo
couso@uniovi.es

**Abstract.** The notion of preference is reviewed from different perspectives, including the Imprecise Probabilities' approach. Formal connections between different streams of the literature are provided, and new definitions are proposed.

**Keywords:** Imprecise probabilities, stochastic orderings, Walley's desirability, fuzzy rankings, possibility theory.

## 1 Introduction

The problem of finding appropriate rankings to evaluate and/or sort different alternatives appears in a variety of disciplines such as Economics ([2]), Political Sciences ([5]) or Artificial Intelligence ([16]), among many others. We can usually identify four main ingredients in a decision making problem, namely:

1. The set of alternatives (also referred to as "gambles" ([36]), "actions" ([4,32]),"options" ([3,24]) or "candidates" ([5,39]).
2. The "set of states of nature" –also called "criteria" in multi-criteria decision problems ([4,20,25,32]), "experts" ([13]) or "individuals" in group decision making ([22]), or "voters" in political decision making ([5,26])–.
3. An evaluation on every alternative and/or a preference ordering between different alternatives for each state of nature.
4. A merging method in order to combine evaluations or preference relations for different states of nature, that allows us to select the best (or a set of non dominated) alternative(s).

Different streams of the literature make different assumptions about the two last ingredient. With respect to the preference ordering over the set of alternatives, for each particular state of the nature, we face at least two different approaches:

* The author is very thankful to the Executive Directors and the Organizers for their kind invitation to participate in the conference. The research in this work has been supported by the Spanish grant TIN2011-24302. This financial support is gratefully acknowledged.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 1–9, 2014.

3a. In the first one, a quantitative [12,10,20,35,36] or a qualitative (see [31], for instance) assessment is associated to every single alternative. In the case of quantitative assessments, the usual ordering between numbers can be a natural choice to sort the set of alternatives, although it is not the only option ([21]). Recently, (fuzzy) interval-valued assessments instead of crisp numerical values have been considered in some problems where imprecision is involved in this assessment process [1,34]), and therefore specific partial or total pre-orderings between (fuzzy) intervals must be selected in order to sort the different alternatives, for every particular state of nature.

3b. According to the second approach, preference statements over the set of alternatives are made, instead of considering value functions providing single assessments. Those preference statements can be given by means of graphical representations, for instance, and may lead to a partial ordering over the set of alternatives, for each state of nature or criterion (see [18,25], for instance.)

Regarding the combination of our preference orderings under the different states of nature into a single (partial) ordering, we can distinguish at least three approaches:

4a. The expert initial information is assessed by means of comparative preference statements between options. A (family of) probability measures(s) over the states of nature is derived from it. This is the "behavioral" approach initiated by Savage in a precise probabilities environment ([11,33]) and followed afterwards in the general theory of imprecise probabilities (see [8,36]).

4b. A (family of ) probability measure(s) over the class of states of nature is initially considered. A preference relation on the set of alternatives ("random variables" or "gambles") is derived from it. This is the approach considered for instance in [21,9,12], for the case of precise probabilities and in [14,29,34,35]. for the case of partially known probabilities or weights. Such a preference relation, derived from the initial (sometimes partial) information about the weights of the different states of nature (or "criteria", in multi-criteria decision), is closely related to the notion of "almost preference" in Walley's framework. This approach will be reviewed in Subsection 3.2 ("credal set approach").

4c. This approach usually appears in multi-criteria decision problems. Traditionally, a weight function, formally equivalent to a probability measure, is defined over the (finite) set of criteria, in order to determine the relative importance of each of them. The global score of an alternative is therefore calculated according to the weighted mean, that can be seen as a (discrete) Lebesgue expectation. Lately ([27]) some authors have generalized this approach in order to include the possibility of interaction between different criteria. In order to do so, the weights associated to some families of criteria do not necessarily coincide with the sum of the weights of each of them. This kind of complex information can be represented by means of a non additive set-function defined over the set of criteria. According to this approach, the global score of each alternative may be determined by the Choquet integral of

the partial scores with respect to such a set function. We will briefly review of this approach in Subsection 3.3 ("aggregation operators-based approach").

Once we have described the four usual ingredients in decision making problems, we need to take into account an additional issue. The decision-maker is usually interested by two questions: on the one hand, (s)he may wish to know what decisions are preferred to others; on the other hand, (s)he may be interested in determining whether single alternatives are "satisfactory" ([28]) or "desirable" ([36]) or not. P. Walley ([36]) established a formal link between both issues.

The goals in this paper are threefold:

- First, we aim to highlight what are the formal connections and commonalities between different combination methods (classical stochastic orderings, Walley's preferences, etc.) from the literature.
- Second, we will show Walley's partial orderings as a generalization of the expected utility criterion ([33]), and we will explore other possible generalizations of stochastic orderings within the Imprecise Probabilities setting.
- Finally, we will face the problem of ranking fuzzy intervals from the perspective of Imprecise Probabilities, assuming that each fuzzy interval characterizes a possibility distribution, considered as an upper probability.

Sections 2 and 3 deal with the two first goals, and Section 4, with the third one.

## 2 Preference Modeling in a Probabilistic Setting: Stochastic Orderings

Let us first consider a probability space $(\Omega, \mathcal{F}, P)$ determining the "weights" of the different states of the nature. Let us consider a pair of random variables (alternatives) defined on $\Omega$, $X, Y : \Omega \to \mathbb{R}$. This section reviews three well known stochastic orderings in the literature:

SO1. *Dominance in the sense of expected utility* [33].- Given an increasing function $u : \mathbb{R} \to \mathbb{R}$, $X$ dominates $Y$ wrt $u$ if $E_P(u(X)) \geq E_P(u(Y))$. We will denote it $X \geq_u Y$. A special case is *Dominance in Expectation*: $X$ dominates $Y$ if $E_P(X) \geq E_P(Y)$. This relation represents the particular case of the previous one, when the utility function $u$ is the identity function $u(x) = x$, $\forall\, x \in \mathbb{R}$.

SO2. *Statistical preference* [9].- $X$ is statistically preferred to $Y$ if $P(X > Y) + 0.5 \cdot P(X = Y) \geq 0.5$. We will denote it $X \geq_{SP} Y$.

SO3. *First order stochastic dominance* [21].- $X$ dominates $Y$ if $P(X > a) \geq P(Y > a), \forall\, a \in \mathbb{R}$. It is well known that $X \geq_{1st} Y$ if and only if $X \geq_u Y$, for all increasing utility functions $u : \mathbb{R} \to \mathbb{R}$. We will denote it $X \geq_{1st} Y$.

According to [7], all the above stochastic orderings can be put into a common formulation. In fact, we can express each of them as follows:

$$X \text{ is preferred to } Y \text{ if and only if } E_P[f(X,Y)] \geq E_P[f(Y,X)], \tag{1}$$

$$\text{or equivalently, if } E_P[f(X,Y) - f(Y,X)] \geq 0, \tag{2}$$

for a certain function $f : \mathbb{R}^2 \to \mathbb{R}$, increasing in the first component and decreasing in the second one. Below, we provide the specific expression of $f(\cdot,\cdot)$ for each of the above orderings:

- *Dominance in the sense of expected utility.-* Let us consider the function $f_u(x,y) = u(x), \forall\, x \in \mathbb{R}$ (which is a constant wrt the second argument). We can easily check that $X \geq_u Y$ if and only if $E_P[f_u(X,Y)] \geq E_P[f_u(Y,X)]$.
- *Statistical preference.-* Let us consider the mapping $f(x,y) = \text{sgn}(x-y)$, where sgn denotes the "sign" function taking the values $1, 0$ or $-1$, depending on the sign of the argument. It is checked in [7,8] that $X \geq_{SP} Y$ if and only if $E[f(X,Y)] \geq E[f(Y,X)]$.
- *First stochastic dominance.-* Let us now consider the family of functions $f^a(x,y) = 1_{x>a} - 1_{y>a},\ \forall\, a \in \mathbb{R}$. $X$ dominates $Y$ if and only if $E[f^a(X,Y)] \geq E[f^a(Y,X)],\ \forall\, a \in \mathbb{R}$. Equivalently, and according to the formal existing relation between dominance in the sense of expected utility and first stochastic dominance, we can say that $X$ dominates $Y$ if and only if $E[f_u(X,Y)] \geq E[f_u(Y,X)]$, for every (increasing) utility function $u : \mathbb{R} \to \mathbb{R}$.

## 3 Partial Orderings in Imprecise Probabilities

### 3.1 The Behavioral Approach

Peter Walley [37] establishes a list of axioms of coherence for preference relations between gambles. A preference relation $\succeq$ defined on a linear space of gambles[1] $\mathcal{K}$ is said to be *coherent* when it satisfies the following properties:

P1. Not $X \succeq X$.
P2. If $X(\omega) \geq Y(\omega),\ \forall\, \omega \in \Omega$ and $X(\omega) > Y(\omega)$, for some $\omega \in \Omega$, then $X \succeq Y$.
P3. If $X \succeq Y$ and $c > 0$ then $cX \succeq cY$.
P4. If $X \succeq Y$ and $Y \succeq Z$ then $X \succeq Z$.
P5. $X \succeq Y$ if and only if $X - Y \succeq 0$.

He also establishes a duality between the notions of *desirability* and *preference*: $X$ is desirable if and only if it is preferred to the null gamble (the gamble that provides no reward, no matter the state of the nature, $\omega \in \Omega$.) Conversely, $X$ is preferred to $Y$ when their difference $X - Y$ is a desirable gamble. In such a case, one is willing to give up $Y$ in return for $X$. Any coherent preference relation

[1] A gamble $X : \Omega \to \mathbb{R}$ is a bounded mapping defined on the space of states of nature. If you were to accept gamble $X$ and $\omega$ turned to be true, then you would gain $X(\omega)$. (This reward can be negative, and then it will represent a loss.)

determines a *lower prevision*, $\underline{P}$, representing the supremum of acceptable buying prices for gambles, that is formally defined as follows:

$$\underline{P}(X) = \sup\{c \in \mathbb{R} : X \succeq c\}, \ \forall X.$$

The lower prevision $\underline{P}$, determines the credal set[2], $M(\underline{P})$, formed by the set of linear previsions that dominate $\underline{P}$: $M(\underline{P}) = \{P : P(X) \geq \underline{P}(X), \ \forall X\}$. According to the last formula, it can be interpreted as the minimum of the expectation operators associated to a convex family of (finitely additive) probabilities. Furthermore, the following implications hold for any pair of gambles, $X$, $Y$:

$$\underline{E}(X - Y) > 0 \Rightarrow X \text{ is preferred to } Y \Rightarrow \underline{E}(X - Y) \geq 0. \tag{3}$$

According to Equation 3, Walley's preference clearly generalizes the "dominance in expectation" criterion reviewed in Section 2.

We have recently explored ([8]) the generalization of statistical preference, defining the preference of $X$ over $Y$ as the desirability of the sign of their difference, $\text{sgn}(X - Y)$. Under this criterion, we just take into account whether the consequent $X(\omega)$ is greater than (or preferred to) $Y(\omega)$ or not, but not the magnitude of their difference. Thus, it does not require the existence of a numerical scale, unlike Walley's criterion. The following equation reminds Eq. 3, where the lower expectation has been replaced by the median:

$$\underline{\text{Me}}(X - Y) > 0 \Rightarrow X \ \text{ is } \ \text{signed} \ \ \text{preferred} \ \ \text{to } Y \Rightarrow \underline{\text{Me}}(X - Y) \geq 0.$$

Taking into account the last section, and some of the ideas suggested in [7], the notion of first stochastic dominance could be easily generalized to the imprecise probabilities setting, if we consider that $X$ is preferred to $Y$ whenever the gamble $1_{X>a}$ is preferred to $1_{Y>a}$, for every $a \in \mathbb{R}$.

### 3.2 The Credal Set-Based Approach

Lately, different authors have independently generalized some stochastic orderings to the case where our imprecise information about the underlying probability distribution, $P$, over the set of states of nature is determined by means of a set of probability measures $\mathcal{P}$. Taking into account the general formulation of stochastic orderings proposed in Section 2, many of those new definitions can be seen from a general perspective, where the values $E_P[f(X,Y)]$, $E_P[f(Y,X)]$ and $E_P[f(X,Y)] - E_P[f(Y,X)]$ are replaced by their respective sets of admissible values. More specifically, they can be seen as generalizations of Eq. 1 and 2, where the inequality $\geq$ is replaced by a particular (sometimes partial) preorder between intervals (or, more generally, between arbitrary sets of numbers). In this respect, the criteria considered in [35] generalize the "dominance in expectation"

[2] A credal set is a convex and closed family of linear previsions, i.e., of linear functionals $P : \mathcal{L} \to \mathbb{R}$ satisfying the constraint $P(1) = 1$. The last concept generalizes the notion of expectation operator to the case of non necessarily $\sigma-$additive probabilities.

criterion (SO1). On the other hand, the criterion considered in [34] generalizes the notion of statistical preference (SO2), to the case where the usual ordering between numbers is replaced by an interval ordering ([19]). As a side remark, let us notice that there exist a formal connection between this last criterion and the notion of *fuzzy preference* ([30]). In fact, it can be easily checked that the mapping $f$ defined as $f(X,Y) = \overline{P}(X,Y)$ is a *connected* [3] or *complete* fuzzy preference, since it satisfies the general axioms of fuzzy preferences and the inequality $f(X,Y) + f(Y,X) \geq 1$. If, in addition, $\overline{P}$ is a possibility measure, then the fuzzy preference relation derived from it is *strongly connected*. Finally, the four generalizations of "first stochastic dominance" (SO3) proposed in [14] and the six ones proposed in [29] also follow the general procedure described in this section. A more detailed explanation about how the criteria reviewed in this section fit this general formulation is provided in [7].

### 3.3 The Aggregation Operators-Based Approach

The weighted mean is a very often used aggregation criterion in multicriteria decision problems, and it is formally equivalent to "dominance in expectation", where the probability masses play the role of "weights of importance" of the different criteria instead of being interpreted in terms of stochastic uncertainty. During the last decades (see [27] and references therein), generalizations of the weighted mean operator, like the discrete Choquet integral (that includes OWA operators [38] as particular cases) have been considered. Within this more general setting, the degree of importance associated to a particular set of criteria is not forced to coincide with the sum of the weights assigned to the particular criteria included in the set, allowing the possibility of modeling the effect of interaction between different criteria. The discrete Choquet integral wrt the non-additive monotone set-function, $\mu$, that assigns the weight $\mu(C)$ to each particular set of criteria, $C$, seems to be the natural generalization of the weighted mean under this general approach. If the non-additive set function satisfies some additional properties, like *submodularity*, for instance, the Choquet integral plays the role of a lower prevision or a lower expectation. In those cases, the resulting aggregation method is formally linked to the preference criterion defined by Walley reviewed in Section 3.1, as well as to the generalizations of the "dominance in expectation" criteria considered in Section 3.2.

Labreuche et al. [25] also interpret the "weights" as relative degrees of importance of criteria. In their paper, partial preference relations between the options, instead of single evaluations are considered for each particular criterion (see 3a). Those preference relations are combined afterwards according to "statistical preference" (SO2). Weights on sets of criteria are not directly provided. Instead, preference relations between indicator functions over the set of criteria are given, and the set of feasible probabilities (weight vectors) is derived from them. So the approach in this paper is very much connected to the procedure described in (4a). It would be interesting to explore the formal connection between this approach and the preference criterion proposed in [8].

## 4 An Imprecise Probabilities' Approach to the Notion of Fuzzy Ranking

According to the possibilistic interpretation of fuzzy sets, the problem of ranking fuzzy numbers can be seen from an Imprecise Probabilities' perspective ([6,17]. In fact, a pair of fuzzy sets can be seen as an incomplete description of a joint probability measure $P_{(X,Y)}$, and therefore, any of the criteria reviewed in Sections 2 and 3 could be applied. For instance, Detyniecki et al. [15] apply statistical preference to the joint probability distribution induced by a pair of independent random variables whose respective density functions are proportional to the respective fuzzy sets membership functions, considered as possibility distributions. Sánchez et al. ([34]) also generalize statistical preference, but this time, they consider the whole set of probability measures dominated by a joint upper probability determined by both possibility distributions. The criterion of dominance of expectation has been also generalized in the recent literature (see [10] for instance). A deep analysis studying well-known fuzzy rankings from the perspective of Imprecise Probabilities has been developed in [6].

## References

1. Aiche, F., Dubois, D.: An extension of stochastic dominance to fuzzy random variables. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) IPMU 2010. LNCS, vol. 6178, pp. 159–168. Springer, Heidelberg (2010)
2. Arrow, K.J.: Social Choice and Individual Values, 2nd edn. Wiley, New York (1963)
3. Barret, R., Salles, M.: Social choice with fuzzy preferences. In: Handbook of Social Choice and Welfare, pp. 367–389. Elsevier (2004)
4. Chiclana, F., Herrera, F.: Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations. Fuzzy Sets Syst. 97, 33–48 (1998)
5. Condorcet, M.: Essay sur le application de l'analyse à la probabilité des decisions rendues à la pluralité des voix. Imprimirie Royale, Paris (1785)
6. Couso, I., Destercke, I.: Ranking of fuzzy numbers seen through the imprecise probabilistic lense. In: Eurofuse 2013 (2013)
7. Couso, I., Dubois, D.: An imprecise probability approach to joint extensions of stochastic and interval orderings. In: Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R.R. (eds.) IPMU 2012, Part III. CCIS, vol. 299, pp. 388–399. Springer, Heidelberg (2012)
8. Couso, I., Sánchez, L.: The behavioral meaning of the median. In: Borgelt, C., González-Rodríguez, G., Trutschnig, W., Lubiano, M.A., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (eds.) Combining Soft Computing and Statistical Methods in Data Analysis. AISC, vol. 77, pp. 115–122. Springer, Heidelberg (2010)
9. David, H.: The method of paired comparisons, Griffin's Statistical Monographs & Courses, vol. 12. Charles Griffin & D. Ltd., London (1963)
10. De Cooman, G.: Further thoughts on possibilistic previsions: A rejoinder. Fuzzy Sets Syst. 153, 375–385 (2005)
11. De Finetti, B.: Theory of Probability, vol. 1 & 2. John Wiley & Sons (1974-1975)
12. De Schuymer, B., De Meyer, H., De Baets, B., Jenei, S.: On the cycle-transitivity of the dice model. Theory and Decision 54, 261–285 (2003)

13. Delgado, M., Herrera, F., Herrera-Viedma, E., Martínez, L.: Combining numerical and linguistic information in group decision making. Inform. Sciences 107, 177–194 (1998)
14. Denœux, T.: Extending stochastic ordering to belief functions on the real line. Inform. Sciences 179, 1362–1376 (2009)
15. Detyniecki, M., Yager, R.R., Bouchon-Meunier, B.: A context-dependent method for ordering fuzzy numbers using probabilities. Inform. Science 138, 237–255 (2001)
16. Domshlak, C., Hüllermeier, E., Kaci, S., Prade, H.: Preferences in AI: An overview. Artificial Intelligence 175, 1037–1052 (2011)
17. Dubois, D.: The role of fuzzy sets in decision sciences: Old techniques and new directions. Fuzzy Sets and Systems 184, 3–28 (2011)
18. Dubois, D., Fargier, H., Prade, H.: Decision making under ordinal preferences and comparative uncertainty. In: Proceedings of International Conference on Uncertainty in Artificial Intelligence (UAI 1997), pp. 157–164 (1997)
19. Fishburn, P.C.: Interval Orderings and interval graphs. Wiley, New-York (1985)
20. Grabisch, M., Labreuche, C.: A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. Ann. Oper. Res. 175, 247–286 (2010)
21. Hadar, J., Russell, W.: Rules for Ordering Uncertain Prospects. American Economic Review 59, 25–34 (1969)
22. Herrera, F., Herrera-Viedma, E., Verdegay, J.L.: A model of consensus in group decision making under linguistic assessments. Fuzzy Sets Syst. 78, 73–87 (1996)
23. Kaci, S., Labreuche, C.: Argumentation framework with fuzzy preference relations. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) IPMU 2010. LNCS, vol. 6178, pp. 554–563. Springer, Heidelberg (2010)
24. Kacprzyk, J.: Group decision making with a fuzzy linguistic majority. Fuzzy Sets Syst. 18, 105–118 (1986)
25. Labreuche, C., Maudet, N., Ouerdane, W.: Justifying Dominating Options when Preferential Information is Incomplete. In: ECAI 2012, pp. 486–491 (2012)
26. Ladha, K.K.: The Condorcet Jury Theorem, Free Speech, and Correlated Votes. American Journal of Political Science 36, 617–634 (1992)
27. Marichal, J.L.: An Axiomatic Approach of the Discrete Choquet Integral as a Tool to Aggregate Interacting Criteria. IEEE Trans. Fuzzy Syst. 8, 800–807 (2000)
28. Marichal, J.L.: Aggregation Functions for Decision Making. In: Bouysou, et al. (eds.) Decision-making Process. Concepts and Methods. Wiley (2009)
29. Montes, I., Miranda, E., Montes, S.: Stochastic dominance with imprecise information. Computational Statistics & Data Analysis 71, 868–886 (2014)
30. Orlovsky, S.: Decision-making with a fuzzy preference relation. Fuzzy Sets Syst. 1, 155–168 (1978)
31. Rico, A., Grabisch, M., Labreuche, C., Chateauneuf, A.: Preference modeling on totally ordered sets by the Sugeno integral. Discrete Appl. Math. 147, 113–124 (2005)
32. Roubens, M.: Preference relations on actions and criteria in multicriteria decision making. Eur. J. Oper. Res. 10, 51–55 (1982)
33. Savage, L.J.: The Foundations of Statistics. Wiley (1954)
34. Sánchez, L., Couso, I., Casillas, J.: Genetic Learning of Fuzzy Rules based on Low Quality Data. Fuzzy Sets and Systems 160, 2524–2552 (2009)
35. Troffaes, M.: Decision making under uncertainty using imprecise probabilities. Int. Journal Appr. Reas. 45, 17–29 (2007)

36. Walley, P.: Statistical reasoning with imprecise probabilities. Chapman and Hall, London (1991)
37. Walley, P.: Towards a unified theory of imprecise probability. Int. J. Appr. Reas. 24, 125–148 (2000)
38. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Trans. Syst., Man, Cybern. 18, 183–190 (1988)
39. Young, H.P.: Condorcet's Theory of Voting. American Political Science Review 82, 1231–1244 (1988)

# Unifying Logic and Probability: A New Dawn for AI?

Stuart Russell

[1] University of California, Berkeley CA 94720, USA
russell@cs.berkeley.edu
http://www.cs.berkeley.edu/~russell
[2] Université Pierre et Marie Curie, Laboratoire d'informatique de Paris 6,
4 place Jussieu, 75252 Paris Cedex 05, France

**Abstract.** Logic and probability theory are two of the most important branches of mathematics and each has played a significant role in artificial intelligence (AI) research. Beginning with Leibniz, scholars have attempted to unify logic and probability. For "classical" AI, based largely on first-order logic, the purpose of such a unification is to handle uncertainty and facilitate learning from real data; for "modern" AI, based largely on probability theory, the purpose is to acquire formal languages with sufficient expressive power to handle complex domains and incorporate prior knowledge. This paper provides a brief summary of an invited talk describing efforts in these directions, focusing in particular on *open-universe* probability models that allow for uncertainty about the existence and identity of objects.

**Keywords:** first-order logic, probability, probabilistic programming, Bayesian logic, machine learning.

## 1 Introduction

From its earliest days, AI adopted the idea of *declarative* system reasoning over explicitly represented knowledge with a general inference engine. Such systems require a formal language to express knowledge about the real world; and *the real world has things in it.* For this reason, in 1958, McCarthy [16] proposed first-order logic—the mathematics of objects and relations—as the foundation for what we now call "classical AI."

The key benefit of first-order logic is its expressive power, which leads to concise—and hence easily learnable—models. For example, the rules of chess occupy $10^0$ pages in first-order logic, $10^5$ pages in propositional logic, and $10^{38}$ pages in the language of finite automata. The power comes from separating predicates from their arguments and quantifying over those arguments: so one can write rules about $On(p, c, x, y, t)$ (piece $p$ of color $c$ is on square $x, y$ at move $t$) without having to fill in each specific value for $c$, $p$, $x$, $y$, and $t$.

A second research tradition, sometimes called "modern AI," developed around another important property of the real world: *pervasive uncertainty* about both

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 10–14, 2014.

its state and its dynamics. Modern AI is based on probability theory, which provides principled methods for learning and making predictions from observations. The key advance underlying modern AI was the development of *Bayesian networks* [22] and the related family of undirected *graphical models* [6]. Bayes nets provided a formal language for probability models and enabled rapid advances in machine learning, vision, natural language understanding, and knowledge-based systems. The expressive power of Bayes nets is, however, limited. They assume a fixed set of *variables*, each of which can take on a value from a fixed *range*; thus, they are a *propositional* formalism, like Boolean circuits. The rules of chess and of many other domains are beyond them.

What happened next, of course, is that classical AI researchers noticed the pervasive uncertainty, while modern AI researchers noticed, or remembered, that the world has things in it. Both traditions arrived at the same place: *the world is uncertain* **and** *it has things in it.* To deal with this, we have to *unify logic and probability.*

But how? Even the meaning of such a goal is unclear. Early attempts by Leibniz, Bernoulli, De Morgan, Boole, Peirce, Keynes, Carnap, and Gaifman (surveyed in [8,10]) involved attaching probabilities to logical sentences. This line of work influenced AI research [9,3,14] but has serious shortcomings as a vehicle for representing knowledge. An alternative approach, arising from both branches of AI and from statistics, draws on the compositional semantics of Bayes nets. Some tools use programming constructs to build very large Beys nets with repeated structure [7,4,15], while others adopt the syntactic and semantic devices of logic (composable function symbols, logical variables, quantifiers) to create declarative, first-order probabilistic languages [5,23,25,12,11].

Despite their successes, these approaches miss an important consequence of uncertainty in a world of things: there will be *uncertainty about what things are in the world.* Real objects seldom wear unique identifiers or preannounce their existence like the cast of a play. In the case of vision, for example, the existence of objects must be *inferred* from raw data (pixels) that contain no explicit object references at all. If, however, one has a probabilistic model of the ways in which worlds can be composed of objects and of how objects cause pixel values, then inference can propose the existence of objects given only pixel values as evidence. Similar arguments apply to areas such as natural language understanding, web mining, and computer security.

The difference between knowing all the objects in advance and inferring their existence and identity from observation corresponds to an important but often overlooked distinction between *closed-universe* languages such as SQL and logic programs and *open-universe* languages such as full first-order logic.

This distinction is best understood in terms of the *possible worlds* under each type of semantics. Figure 1(a) shows a simple example with two constants and one binary predicate. Notice that first-order logic is an open-universe language: even though there are two constant symbols, the possible worlds allow for 1, 2, or indeed arbitrarily many objects. A closed-universe language enforces additional assumptions that restrict the set of possible worlds:

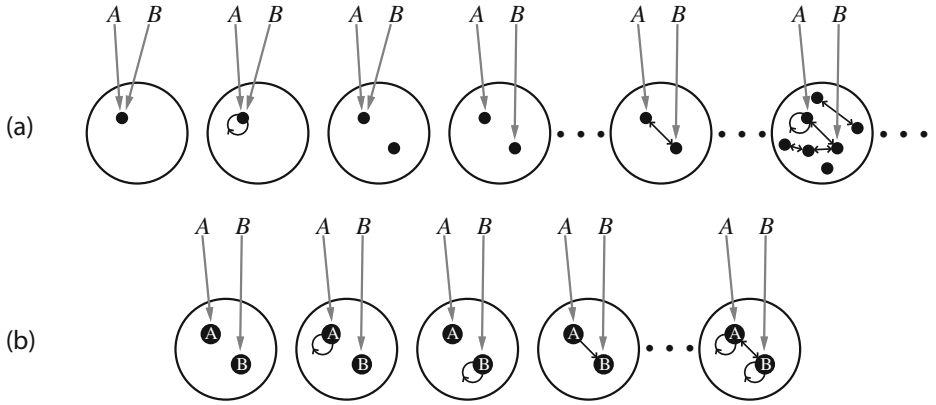



**Fig. 1.** (a) Some of the first-order possible worlds for a language with two constant symbols, $A$ and $B$, and one binary predicate. Arrows indicate the interpretation of each constant symbol and the relations between objects. (b) The analogous figure under closed-universe semantics.

– The *unique names* assumption requires that distinct terms must refer to distinct objects.
– The *domain closure* assumption requires that there are no objects other than those named by terms.

These two assumptions mean that every possible world contains the same objects, which are in one-to-one correspondence with the ground terms of the language (see Figure 1(b)).[1]

A formal probability model must specify the probability of every possible world given the vocabulary (predicates, functions, constants) of the model's syntactic representation. Obviously, the set of worlds under open-universe semantics is larger and more heterogeneous, which makes the task of defining open-universe probability models more challenging. The core part of the talk is concerned with a first-order, open-universe probabilistic language called *Bayesian logic* or BLOG [18,19]. BLOG was developed primarily as the PhD thesis research of Brian Milch [17]. The key results derived for BLOG are the following:

– Every well-formed BLOG model specifies a well-defined probability distribution over the possible worlds constructed from the vocabulary of the model.
– There exist Monte Carlo algorithms that provably converge (subject to technical conditions on the conditional distributions of the model) to the correct posterior probability for any first-order query for any well-formed BLOG model [20,1].

[1] The difference between open and closed universes can also be illustrated with a common-sense example. Suppose a system knows just two sentences, $Father(William) = Bill$ and $Father(Junior) = Bill$. How many children does Bill have? Under closed-universe semantics—e.g., in a database system—he has exactly 2; under open-universe semantics, between 1 and $\infty$.

The generic algorithms (importance sampling and MCMC applied to a dynamically constructed ground representation) are often too slow for practical use on large models. Several avenues are being pursued for speeding up inference, including special-purpose block samplers for variables constrained by deterministic relationships [13], static analysis to identify submodels amenable to efficient inference, lifted inference to avoid grounding by manipulating symbolic distributions over large sets of objects [24,26], and compiler techniques to generate model-specific inference code.

More than two dozen BLOG models have been developed, covering a wide variety of standard machine learning models as well as applications including citation matching [21] and global seismic monitoring for the Comprehensive Nuclear Test-Ban Treaty [2].

## 2 Prospects

These are very early days in the process of unifying logic and probability. We need much more experience in developing models for a wide range of applications. Undoubtedly there are new modeling idioms, programming constructs, and inference algorithms to discover.

The development of Bayes nets in the late 1980s connected machine learning to statistics and reconnected (modern) AI with vision and language. It is possible that first-order probabilistic languages, which have both Bayes nets and first-order logic as special cases, can serve a similar, but more inclusive, unifying role.

**Acknowledgments.** The author is supported by the *Chaire Blaise Pascal*, funded by the l'État et la Région Île de France and administered by the Fondation de l'École Normale Supérieure. The research is also supported by DARPA contract FA8750-14-C-0011 for the APPRIL project under the PPAML program and previously by DARPA contract FA8650-11-1-7153 for the OUTBIDS project under the MSEE program.

## References

1. Arora, N., Russell, S., de Salvo Braz, R., Sudderth, E.: Gibbs sampling in open-universe stochastic languages. In: UAI 2010 (2010)
2. Arora, N.S., Russell, S., Sudderth, E.: NET-VISA: Network processing vertically integrated seismic analysis. Bull. Seism. Soc. Amer. 103 (2013)
3. Bacchus, F.: Representing and Reasoning with Probabilistic Knowledge. MIT Press (1990)
4. Bessière, P., Mazer, E., Ahuactzin, J.M., Mekhnacha, K.: Bayesian programming. CRC (2013)
5. Breese, J.S.: Construction of belief and decision networks. Computational Intelligence 8, 624–647 (1992)
6. Darroch, J.N., Lauritzen, S.L., Speed, T.P.: Markov fields and log-linear interaction models for contingency tables. The Annals of Statistics 8(3), 522–539 (1980)

7. Gilks, W.R., Thomas, A., Spiegelhalter, D.J.: A language and program for complex Bayesian modelling. The Statistician 43, 169–178 (1994)
8. Hailperin, T.: Probability logic. Notre Dame J. Formal Logic 25(3), 198–212 (1984)
9. Halpern, J.Y.: An analysis of first-order logics of probability. AIJ 46(3), 311–350 (1990)
10. Howson, C.: Probability and logic. J. Applied Logic 1(3-4), 151–165 (2003)
11. Kersting, K., De Raedt, L.: Bayesian logic programs. In: ILP 2000 (2000)
12. Koller, D., Pfeffer, A.: Probabilistic frame-based systems. In: AAAI 1998 (1998)
13. Li, L., Ramsundar, B., Russell, S.: Dynamic scaled sampling for deterministic constraints. In: AI/Stats 2013 (2013)
14. Lukasiewicz, T.: Probabilistic logic programming. In: ECAI (1998)
15. McCallum, A., Schultz, K., Singh, S.: FACTORIE: Probabilistic programming via imperatively defined factor graphs. NIPS 22 (2010)
16. McCarthy, J.: Programs with common sense. In: Proc. Symposium on Mechanisation of Thought Processes. Her Majesty's Stationery Office (1958)
17. Milch, B.: Probabilistic Models with Unknown Objects. Ph.D. thesis, UC Berkeley (2006)
18. Milch, B., Marthi, B., Sontag, D., Russell, S.J., Ong, D., Kolobov, A.: BLOG: Probabilistic models with unknown objects. In: IJCAI 2005 (2005)
19. Milch, B., Russell, S.: Extending Bayesian networks to the open-universe case. In: Dechter, R., Geffner, H., Halpern, J. (eds.) Heuristics, Probability and Causality: A Tribute to Judea Pearl. College Publications (2010)
20. Milch, B., Russell, S.J.: General-purpose MCMC inference over relational structures. In: UAI 2006 (2006)
21. Pasula, H., Marthi, B., Milch, B., Russell, S.J., Shpitser, I.: Identity uncertainty and citation matching. NIPS 15 (2003)
22. Pearl, J.: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann (1988)
23. Poole, D.: Probabilistic Horn abduction and Bayesian networks. AIJ 64, 81–129 (1993)
24. Poole, D.: First-order probabilistic inference. In: IJCAI 2003 (2003)
25. Sato, T., Kameya, Y.: PRISM: A symbolic statistical modeling language. In: IJCAI 1997 (1997)
26. Van den Broeck, G.: Lifted Inference and Learning in Statistical Relational Models. Ph.D. thesis, Katholieke Universiteit Leuven (2013)

# Two Procedures for Analyzing the Reliability of Open Government Data

Davide Ceolin[1], Luc Moreau[2], Kieron O'Hara[2], Wan Fokkink[1], Willem Robert van Hage[4], Valentina Maccatrozzo[1], Alistair Sackley[3], Guus Schreiber[1], and Nigel Shadbolt[2]

[1] VU University Amsterdam, The Netherlands
{d.ceolin,w.j.fokkink,v.maccatrozzo,guus.schreiber}@vu.nl
[2] University of Southampton, United Kingdom
{l.moreau,kmo,nrs}@ecs.soton.ac.uk
[3] Hampshire County Council, Winchester, United Kingdom
{alistair.sackley}@hants.gov.uk
[4] SynerScope B.V., Eindhoven, The Netherlands
{willem.van.hage}@synerscope.com

**Abstract.** Open Government Data often contain information that, in more or less detail, regard private citizens. For this reason, before publishing them, public authorities manipulate data to remove any sensitive information while trying to preserve their reliability. This paper addresses the lack of tools aimed at measuring the reliability of these data. We present two procedures for the assessment of the Open Government Data reliability, one based on a comparison between open and closed data, and the other based on analysis of open data only. We evaluate the procedures over data from the data.police.uk website and from the Hampshire Police Constabulary in the United Kingdom. The procedures effectively allow estimating the reliability of open data and, actually, their reliability is high even though they are aggregated and smoothed.

## 1 Introduction

Open Government Data are often sensitive and hence need to be properly processed in order to reduce the amount of personal information exposed. This process consists of aggregation and so-called "smoothing" procedures which introduce some imprecision in the data, to avoid the reconstruction of the identity of a citizen from a piece of data. The value of this data might be affected by such procedures, as they limit the extent to which we can rely on them. Throughout the paper, we will refer to the published Open Government Data as "open data" and to the original data as "closed data".

Open data are exposed in different modalities by different sources. For instance, Crime Reports [5] and `data.police.uk` [15] both publish data about crimes occurring in the UK, but in different format (maps versus CSV files), level of aggregation, smoothing and timeliness (daily versus monthly update). The smoothing process unavoidably introduces some error in the data. There might

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 15–24, 2014.

be other reasons as well for possible reliability differences among these datasets, like the fact that a given dataset is not based on timely data (or, in general, it is generated from questionable data sources) or the fact that an erroneous aggregation process inadvertently introduced some mistakes. For the police, as well as for citizens, it is important to understand how different two sources are, in order to understand how much they can rely on the data they expose. The police, who can access the original, raw data, is interested in measuring the reliability of the open data in order to know how much they can rely on them, e.g., when establishing projects involving citizens. For citizens, it is important to understand the reliability of the different datasets, since that represents a reason why data exposed by authoritative sources may present discrepancies.

Our goal is to cope with the lack of tools and methodologies to actually measure and compare these data. We address this problem by means of a twofold contribution: first, we propose a procedure for computing the reliability of an open dataset, when having at our disposal both the open and the closed data. We apply this procedure on a set of UK police data. We show that the reliability of these data is not highly affected by the smoothing and aggregation procedures applied to them and that this procedure, once properly instantiated, allows guiding the analyzer to the discovery of points of policy changes with regard to open data creation and reliability variations. Second, we show how it is possible to estimate variations in the reliability of the open data by comparing them to each other, when the closed data are not at our disposal. Both procedures aim to measure and compare these datasets from the reliability point of view, and to guide a human analyzer to the discovery of possible critical points (e.g., policy changes, relevant errors) in these datasets. In both cases, *the reliability of an open dataset is measured as the percentage of non-significantly different entries from the corresponding closed dataset.* In the case of the procedure for analyzing open data only, we can only estimate a reliability variation, but not measure it.

The rest of this paper is structured as follows: Section 2 describes related work; Section 3 describes a procedure for determining the reliability of open data given closed data and Section 4 presents a procedure for analyzing open data. In Section 5 we put forward a case study implementation of both procedures. Section 6 provides a final discussion.

## 2 Related Work

This work completes a previous work from the same authors [2], by improving the open data procedure and extending the validation of both procedures. The analysis of open data is increasingly being spread, for instance, by the leading Open Data Institute [14]. Koch-Weser [11] presents a work on the analysis of the reliability of China's Economic Data which, although focused on a different domain, shares with this work the goal to understand the reliability of open data. Tools for the quality estimation of Open Data are being developed (see for instance Talend Open Studio for Data Quality [13] and Data Cleaner [7]). These tools are designed to understand the adherence of data to particular standards,

similar to our goals, but they aim at constituting a proper middleware component of the entire business process of data management and curation. These tools are not limited to monitoring data quality, but they aim also at quantifying the risk and the financial impact of these data, as well as how to intervene in the business process in case of any problem discovered. Our goal is less business-oriented and more targeted, as we aim at developing procedures for measuring and estimating open data reliability. However, this can be seen as a step towards the development of a more comprehensive tool.

A paper of Ceolin et al. [3] shares with the work here presented the statistical approach in modeling categorical Web data and the use of the Wilcoxon signed-rank test (which is a non-parametric hypothesis test that determines whether two probability distributions are significantly different [16]) to measure the reliability of these data. We do not have at our disposal information about the impact of the different processes on the reliability of the resulting data, but in the future we plan to adopt an approach similar to the one of Ebden et al. [6] to understand this a posteriori. Closer to the topic of the case studies analyzed, i.e., the reliability of Police Open Data, this work can be seen as complementary to the one of Cornelli [4], who researches on the reasons citizens have to trust police.

## 3 Procedure for Comparing Closed and Open Data

Closed data are aggregated and smoothed in order to not expose sensitive information when publishing them. Aggregation, that is to present the data at a coarser, higher level than available, preserves the data correctness and reduces their granularity. It is not intended to introduce imprecisions, but a faulty aggregation process or the wrong use of heterogeneous data sources might unexpectedly affect the data reliability. "Smoothing" is an anonymization procedure used especially when aggregation is not sufficient to guarantee anonymity (e.g., in case of data about low-populated areas). By smoothing, authorities voluntarily introduce some small errors in the data so that they remain reliable at a coarse level, but it is not possible (or at least, hard) to reconstruct the details of the single items. We describe a procedure to evaluate the reliability gap existing between open and closed data, if any. The procedure is generic and in Section 5 we propose some possible implementations.

**Select the Relevant Data.** This selection might involve temporal aspects (i.e., only data referring to the relevant period are considered), or their geographical location (select only the data regarding the area of interest). Other constraints and their combination are possible as well.

**Roll Up Categorical Data.** The categories used to classify the categorical data are ordered in hierarchies. Hierarchies are created to define different categories for different refinement levels when presenting categorical data. We cannot increase the refinement of the data categorized in a coarser manner, so we decrease the granularity level of the closed data.

**Roll Up Smoothed Categorical Data.** This step is similar to the previous one, besides the fact that the expected result is not necessarily coincident with the original one since smoothing may affect the data precision.

**Compare the Corresponding Counts.** by using, for instance, the ratio of the correct items over the total amount or the Wilcoxon signed-rank test [16].

## 4 Procedure for Analyzing Open Data

Open data counts may differ from each other with respect to different points of view (absolute differences, data distribution, etc.). We do not know a priori what is the best manner to compare the data counts, so we aggregate several similarity tests performed on pairs of datasets. When analyzing open datasets, we can compare only data about related facts: for instance, the typology of facts can be the same (e.g., crimes in a given area), but the time they refer to differs. The results that we can expect from this kind of analyses are much less detailed and definite than before: since we do not have at our disposal a gold standard, we can not test properly our hypotheses. We estimate points of reliability change using the following procedure, based on the idea that by analyzing the similarity of the datasets using different similarity measures, these changes can emerge.

**Choose One or More Dataset Similarity Scores.** We compute the similarity of two dataset $d_1$ and $d_2$ as: $sim(d_1, d_2) = avg(t_1(d_1, d_2), \ldots, t_n(d_1, d_2))$ where $avg$ computes the average of the results of the similarity scores resulting from the tests $t_i$ on the items (i.e., values) in $d_1$ and $d_2$. We use the Wilcoxon signed-rank test to check if the data counts differ significantly. Other tests are possible as well. The similarity between two dataset is obtained by aggregating these tests using, for instance, a (weighted) arithmetic average. In subjective logic [8], we can treat the tests as "subjective opinions" about the similarity of the two datasets, and merge them using the "fusion" [9] operator to obtain a beta probability distribution describing the probability for each value in $[0, 1]$ to be the correct similarity value.

**Compute the Similarity, with One or More Scores.** Measure the pairwise similarity between each dataset and the one of the following month (crime counts are aggregated on monthly bases).

**Identify Change Points in the Similarity Sequence.** Change points in the similarity sequence are likely to indicate policy changes in the data creation and hence reliability changes resulting from these policy modifications.

**Aggregate All the Evidence of Policy Changes.** Each change point identified in the previous step represents evidence of a policy change. We run more similarity analyses to reduce the risk of error. Since we are dealing with uncertain observations, we also adopt subjective opinions here and we compute a binomial opinion for each dataset.

There can be natural reasons that explain a variation in the data (e.g., a new law) that do not imply a lack of reliability in one of the two datasets. Moreover, a similarity value taken alone may be difficult to interpret: what does it mean that the similarity between two datasets is, e.g., 0.8? So, we focus on similarity trends and not on single values, and we pinpoint variations in such trends, since such variations have a higher chance to indicate a change in the data reliability.

## 5 Case Study - Police Open Data Analyses

We evaluate the procedures that we propose over police data for the Hampshire Constabulary. As open data we adopt the corresponding entries from the `data.police.uk` website, in particular in the interval from April 2011 until December 2012. `data.police.uk` data are released monthly, and aggregated in time within a month and in space to the level of police neighborhoods, that comprise at least eight postal addresses. We focus on the datasets presenting the counts aggregated per police neighborhood because this kind of classification, although not as detailed as the classification per address, allows an easy comparison between entries and reduces the burden of having to geolocate and disambiguate addresses. As closed data, we have at our disposal a series of datasets from the Hampshire Police Constabulary covering the interval from October 2010 until September 2012 and reporting distinct information for each single crime in that period. The two datasets do not perfectly overlap, but we focus mainly on the intersection between the two intervals covered, which still is the largest part of both datasets. We show that the two procedures allow providing similar findings, even though the first procedure is clearly less uncertain than the second one.

### 5.1 Analyzing the Reliability of Police Open Data

We focus on the intersection between the open and the closed data at our disposal (that is, the period from April 2011 until September 2012).The data at our disposal contain: category, date and geographic Cartesian coordinates of the crime. Following the procedure described in Section 3, we compare the distribution of the crime counts among the different categories for each neighborhood using a statistical test and we aggregate the results in a subjective opinion, because we consider the outcomes of the tests as pieces of evidence about the reliability of the open data, and we treat them as error-prone observations.

**Data Preprocessing.** First, we convert the coordinates from the Cartesian system to latitude and longitude using the RGDAL library [1]. Then we look up the postal code that is closest to the point where the crime happened. This step potentially introduces some error in the analyses, because of the approximation in the coordinates conversion and because although looking up the closest postal code to the point that we are analyzing is the best approximation we can make, but it is not always correct. We manually checked some sample items to confirm the robustness of this procedure. Our results show that the impact of these imperfections is limited. It was not possible to compute the postal code of all the points that we had at our disposal, because some data entries were incomplete and some incorrect.

**Select the Relevant Data.** First, we query the MapIt API [12] in order to retrieve the police constabulary each postal code belongs to and discard the crime items not belonging to the Hampshire Constabulary in the closed datasets. Second, we select the open data for the months for which closed data are available. Lastly, we exclude crime counts of categories not shared

between closed and open data. For instance, the "Anti-social behaviour" category is present in the open data but not in the closed data.

**Aggregate the Data.** Also data aggregation is performed in three steps. **Temporal aggregation** is made to group together data about crimes occurring in the same month. **Geographical aggregation** is made to aggregate the data at police neighborhood level. To aggregate the data at neighborhood level, we match zip code and neighborhood by querying the MapIt API [12]. **Categorical aggregation** is performed by aligning the classifications of the crimes in the open and closed datasets, that is, by bringing the closed data at the same, coarse, level as the open data. The categories of open and closed data belong to the same hierarchy, but closed data are classified using fine grained categories, open data using coarser ones.

**Compare the Aggregated Data.** Once the items are brought to the same level of aggregation, the reliability of the open data is measured. The comparison is made at neighborhood level. We **apply a Wilcoxon signed-rank test** to these counts to check (at 95% confidence level) if the two crime counts are significantly different (negative observation) or not (positive observation) and also we **measure the differences between the crime counts**. The results of the comparisons are aggregated using **binomial subjective opinions**, or the equivalent beta distributions. Of these, we use the expected value, $E = \frac{p+1}{p+n+2}$ where $p$ and $n$ are the amounts of non-significantly and significantly different entries respectively. Alternatively, we make use of **arithmetic average**. Given the high number of observations, the difference between the results obtained with the two methods is negligible.

**Results.** We analyze the datasets from April 2011 to September 2012, that is the interval for which we have closed data. We start by comparing the distribution of crime counts per category on the intersection of neighborhoods in the closed and open datasets. We want to check if the distribution of the crime counts in the matching neighborhoods is affected by data manipulation procedures. We present a series of graphs resulting from a series of analyses in R. The closed data at our disposal are quite complete, but they do not match perfectly the open data, as we can see from Fig. 1(a). We apply the Wilcoxon-signed rank test on the crime counts of each matched neighborhood to check if the rank of the crime categories in terms of crime occurrences is preserved. We can see in Fig. 1(c) that the open datasets score high, as to confirm their high reliability, in the overlapping neighborhoods. The error introduced by smoothing may move the geolocation of one crime item to a wrong neighborhood. We extend the open and the closed datasets to have them covering the same neighborhoods: when a neighborhood is not present in one of the two datasets, we assume that it presents zero crimes there. In this way we take the point of view of laymen people, who can deal only with the open data without knowing which are the overlapping neighborhoods. Fig. 1(d) addresses the issue of how the open data are representative of the actual crime distribution in that area. There are at least two trends which, we suspect, correspond to policy changes. One change possibly regards

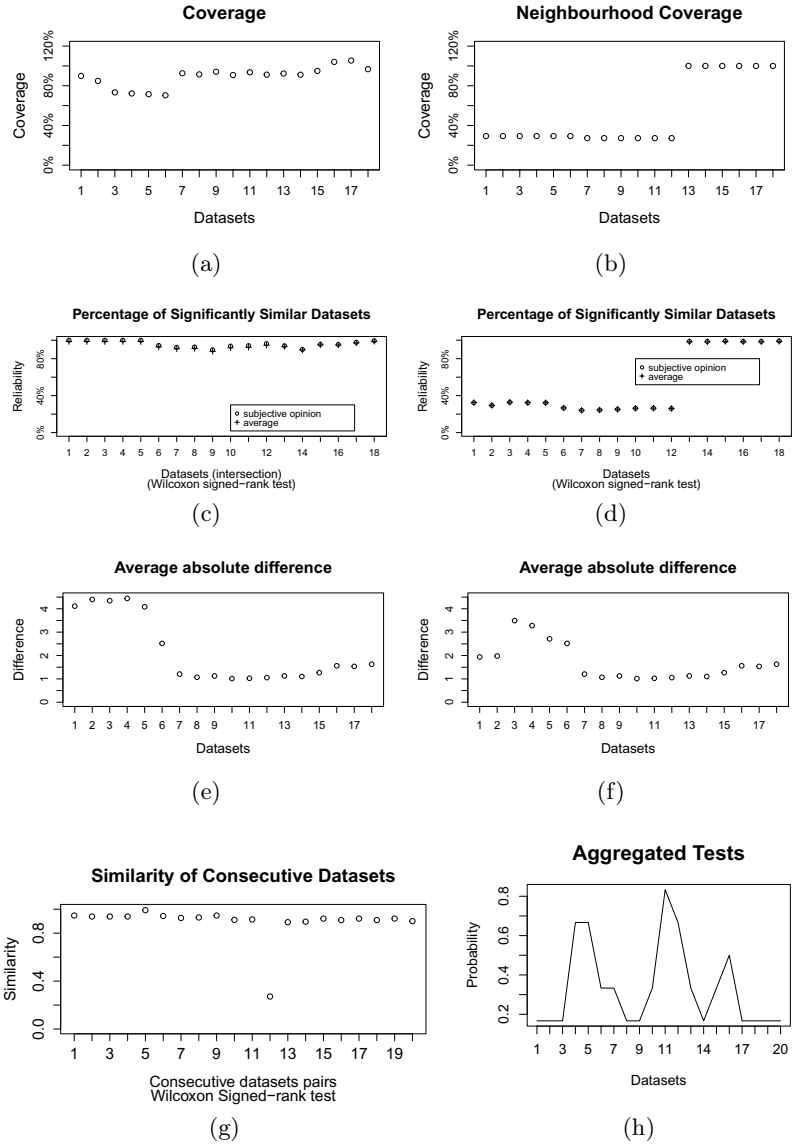



**Fig. 1.** Plots of the open data coverage in terms of counts (Fig. 1(a)) and neighborhoods (Fig. 1(b)). Follow the plots of the closed data analyses using the Wilcoxon signed-rank test (Fig. 1(c) and 1(d)) and count differences (Fig. 1(e) and 1(f)). Finally, we present a plot of the similarity of consequent datasets using again the Wilcoxon signed-rank test (Fig. 1(g)), and plot of the aggregated tests of the similarity of consecutive datasets (Fig. 1(h)).

the smoothing technique adopted, which determines the neighborhood a crime belongs to. Fig. 1(b) shows that initially only about 30% of the neighborhoods were present in both the open and closed datasets, and then this percentage suddenly rose to 100%. This is due to a change in the smoothing algorithm that makes the more recent open data more reliable and explains the "step" shown in Fig. 1(d). Starting from the sixth dataset, the reliability of the extended datasets corresponds to the percentage of matching neighborhoods. In Fig. 1(d) there is also another change point, between the fifth and the sixth dataset. We identify it by checking the absolute errors in the data, averaged per neighborhood and per crime category (see Fig. 1(e)). Here also there are two trends. The first one breaks were approximatively we expected (at the sixth month instead of at the fifth), and these two trends match the trends shown in Fig. 1(a). So we focused on the fifth and sixth datasets (August and September 2011), and three facts emerged.

First, the closed datasets at our disposal do not contain crime counts for the "Drugs" category for July, August and September 2011. We suppose that the closed data at our disposal lack counts for that category in that time interval because that category presents relevant figures in the other months, although it is possible that no drug crime occurred in that period. Still, the procedure correctly identifies a possible anomaly in the data.

Second, in the September 2011 open dataset, the entry relative to the neighborhood "2LW02" presents two very high figures for "Drugs" and "Other.theft", 205 and 319 respectively (the average counts for these categories in the other neighborhoods are 8.44 and 1.68). A similar pattern recurs only in July and September 2012. We suspect that those high counts share the same explanation, and, although we can not verify this with the information at our disposal, the procedure identifies another possible critical point in the data.

Third, a policy change occurred. From September 2011, the set of crime categories was extended to include also {Criminal.damage.and.arson, Shoplifting, Other.theft, Drugs, Public.disorder.and.weapons}. Before, the corresponding crimes were generically classified as "Other.crime". We reclassified the crimes in the first trend belonging to those categories as "Other.crime", and we recomputed the average differences. The error decreases (on average, of 1.56 counts per month, see Fig. 1(f)). In this part, the error is still high because the correct crime classification contains fewer categories than the rest of the datasets, so here the same error weighs more.

Thanks to the procedure proposed, we: (1) discovered two changes in the open data policies that affect data reliability (one about crime classification, one about smoothing); and (2) measured the reliability of these datasets.

### 5.2 Estimating the Reliability of Police Open Data

We analyze open data by applying the following procedure.

**Compute the similarity of the neighborhoods** of consecutive datasets. If more than one similarity measure has been chosen, then aggregate the scores for each neighborhood.

**Aggregate the similarity scores of the neighborhoods** to obtain an overall similarity value. We aggregate using subjective opinions, to take into account also the uncertainty in the sample, that is quite small.

**Look for variations in the series of similarities** that may signal a policy variation, automatically, by means of the *changepoint* package in R [10].

**Aggregate all the evidence per dataset couple** telling whether that couple is a change point or not.

**Results.** We apply the procedure introduced above using different similarity tests (Wilcoxon signed-rank test, absolute differences between counts, etc.). From each test we extrapolate a series of change points, by analyzing the variations in the mean of the cumulative sums (*multiple.mean.cusum* function of the *changepoint* package) and we aggregate them, again by means of an R script and the results are shown in Fig. 1(h). We start from an analysis which is similar to one performed before. We compare, on neighborhood basis, the distribution of the crime counts among the crime categories, and we represent the similarity between two datasets as the percentage of neighborhoods that are statistically similar (using a Wilcoxon signed-rank test). The results of the comparison are reported in Figure 1(g). The datasets are indicated by means of a sequential number (the first circle corresponds to the similarity between the first and the second dataset, and so on). The plot highlights that the twelfth comparison constitutes a change point: before that, the files are highly similar to each other, and likewise after it. But at that point, the similarity trend breaks and starts a new one: that is likely to be a point where the reliability of the datasets diverges. We have found one of the discontinuity points we discovered in Section 5.1 (see Fig. 1(h)). There is also a third peak, less pronounced, but it is not connected to any policy change we are aware of. Also, despite the previous case, we can not say whether a change point indicates the start of an increase or decrease in reliability. Still, these results are useful to facilitate a human analyzer to understand the eventual magnitude of the reliability variation.

## 6 Conclusions

We present two procedures for the computation of the reliability of open data: one based on the comparison between open and closed data, the other one based on the analysis of open data. Both procedures are evaluated using data from the `data.police.uk` website and from the Hampshire Police Constabulary in the United Kingdom. The first procedure effectively allows estimating the reliability of open data, showing also that smoothing procedures preserve a high data reliability, while allowing anonymizing them. Still, we can see an impact of the procedures adopted to produce these data on their reliability, and the most recent policies adopted show a higher ability to preserve data reliability. The second procedure is useful to grasp indications about data reliability and to identify the same critical points detected using the first procedure. The quality of the results achieved with this method is lower than the one achieved with the first method, because it does not allow directly and explicitly estimating the

reliability of the data analyzed. However, the results obtained are useful and retrieved in a semi-automated manner. These two procedures provide additional value to the open data, as they allow enriching the data with information about their reliability: even though these data are already provided by authoritative institutions, these procedures can increase the confidence both of insider specialists (the first procedure, which relies on closed data) and of common citizens (the second procedure, which relies only on open data) who deal with them.

**Acknowledgements.** This work is supported in part under SOCIAM: The Theory and Practice of Social Machines; the SOCIAM Project is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/1.

## References

1. Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R.: RGDAL: Bindings for the Geospatial Data Abstraction Library (2010), https://r-forge.r-project.org/projects/rgdal/
2. Ceolin, D., Moreau, L., O'Hara, K., Schreiber, G., Sackley, A., Fokkink, W., van Hage, W.R., Shadbolt, N.: Reliability Analyses of Open Government Data. In: URSW, pp. 34–39. CEUR-ws.org (2013)
3. Ceolin, D., van Hage, W.R., Fokkink, W., Schreiber, G.: Estimating Uncertainty of Categorical Web Data. In: URSW, pp. 15–26. CEUR-WS.org (2011)
4. Cornelli, R.: Why people trust the police. An empirical study. PhD thesis, Università degli Studi di Trento, International Ph.D. in Criminology (February 13, 2003)
5. CrimeReports. Crimereports (2013), https://www.crimereports.co.uk/
6. Ebden, M., Huynh, T.D., Moreau, L., Ramchurn, S., Roberts, S.: Network analysis on provenance graphs from a crowdsourcing application. In: Groth, P., Frew, J. (eds.) IPAW 2012. LNCS, vol. 7525, pp. 168–182. Springer, Heidelberg (2012)
7. Human Inference. DataCleaner (2013), http://datacleaner.org
8. Jøsang, A.: A logic for uncertain probabilities. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9(3), 279–311 (2001)
9. Jøsang, A.: The consensus operator for combining beliefs. Artificial Intelligence Journal 142, 157–170 (2002)
10. Killick, R., Eckley, I.A.: changepoint: An R Package for Changepoint Analysis (2013), http://www.lancs.ac.uk/~killick/Pub/KillickEckley2011.pdf
11. Koch-Weser, I.N.: The Reliability of China's Economic Data: An Analysis of National Output (2013), http://www.uscc.gov/sites/default/files/Research/TheReliabilityofChina'sEconomicData.pdf
12. Mapit. Mapit (2013), http://mapit.mysociety.orgs
13. Talend. Talend Open Studio for Data Quality (2013), http://www.talend.com/products/data-quality
14. The Open Data Institute. The Open Data Institute (2013), http://www.theodi.org
15. United Kingdom Police Home Office. data.police.uk (2013), http://data.police.uk
16. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bulletin 1(6), 80–83 (1945)

# Classification with Evidential Associative Rules

Ahmed Samet[1,2], Eric Lefèvre[2], and Sadok Ben Yahia[1]

[1] Laboratory of research in Programming, Algorithmic and Heuristic Faculty of Science of Tunis, Tunisia
{ahmed.samet,sadok.benyahia}@fst.rnu.tn
[2] Univ. Lille Nord de France UArtois, EA 3926 LGI2A, F-62400, Béthune, France
eric.lefevre@univ-artois.fr

**Abstract.** Mining database provides valuable information such as frequent patterns and especially associative rules. The associative rules have various applications and assets mainly data classification. The appearance of new and complex data support such as evidential databases has led to redefine new methods to extract pertinent rules. In this paper, we intend to propose a new approach for pertinent rule's extraction on the basis of *confidence* measure redefinition. The confidence measure is based on conditional probability basis and sustains previous works. We also propose a classification approach that combines evidential associative rules within information fusion system. The proposed methods are thoroughly experimented on several constructed evidential databases and showed performance improvement.

**Keywords:** Evidential database, Confidence, Associative classification, Evidential Apriori.

## 1 Introduction

Data mining domain allows extracting pertinent information within databases [1]. The provided information are represented in a set of rules, where each one is associated with a pertinence measure denoted *Confidence*. Among their purposes, those *associative rules* are used for data classification [2,3]. The classification process from those associative rules is denoted *associative classification*. Associative classification offers one of the best classification rate and measure membership [3]. Recently, new databases have appeared proposing data suffering from imperfection. Those types of data fit reality where opinions are no longer represented with Boolean values. In addition, it has added more complexity in their treatment. The imperfection is handled with several theories such as fuzzy [4] and evidential theory [5,6]. In [7], the author introduced a new type of databases that handle both imprecise and uncertain information thanks to the evidential theory. Those types of databases were denoted as the *Evidential database*. The evidential databases were shortly studied from a data mining view [8] and not so much attention was paid to that issue. In literature, two major works [8,9] stand by proposing new measures for itemsets' support. Indeed, in [8], Hewawasam et al. proposed a methodology to estimate itemsets' support

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 25–35, 2014.

and modelize them in a tree representation: *Belief Itemset Tree* (BIT). The BIT representation brings easiness and rapidity for the estimation of the associative rule's confidence. In [9], the authors introduced a new approach for itemset support computing and applied on a Frequent Itemset Maintenance (FIM) problem. Only [8] paid attention to associative classification where the authors introduced *evidential associative rules*. A new measure for rule's confidence was introduced based on conditional belief [6]. In this work, evidential data mining problem is tackled by putting our focus on the associative classification. We highlight problems existing in current measure of evidential rule's confidence which are based on conditional belief. A new confidence measure is proposed based on Bayesian assumption. We also introduce a new associative classification method that reduces the overwhelming number of generated rules. The retained rules are then used for classification purposes and tested on several benchmarks. This paper is organized as follows: in section 2, the main principles of the evidential database are recalled. In section 3, several state of art works on confidence measure are scrutinized and we highlight their limits. In addition, we introduce an alternative confidence measure based on probabilistic definitions. In section 4, we introduce a new method for evidential rule generation. The provided rules are filtrated and combined through a fusion system. The performance of this algorithm is studied in section 5. Finally, we conclude and we sketch issues of future work.

## 2 Evidence Database Concept

An evidential database stores data that could be perfect or imperfect. Uncertainty in such database is expressed via the evidence theory [5,6]. An evidential database, denoted by $\mathcal{EDB}$, with $n$ columns and $d$ lines where each column $i$ $(1 \leq i \leq n)$ has a domain $\theta_i$ of discrete values. Cell of line $j$ and column $i$ contains a normalized BBA as follows:

$$m_{ij} : 2^{\theta_i} \rightarrow [0,1] \quad with$$

$$\begin{cases} m_{ij}(\emptyset) = 0 \\ \sum\limits_{A \subseteq \theta_i} m_{ij}(A) = 1. \end{cases} \tag{1}$$

**Table 1.** Evidential transaction database $\mathcal{EDB}$

| Transaction | Attribute A | Attribute B |
|---|---|---|
| T1 | $m_{11}(A_1) = 0.7$ | $m_{21}(B_1) = 0.4$ |
| | $m_{11}(\theta_A) = 0.3$ | $m_{21}(B_2) = 0.2$ |
| | | $m_{21}(\theta_B) = 0.4$ |
| T2 | $m_{12}(A_2) = 0.3$ | $m_{22}(B_1) = 1$ |
| | $m_{12}(\theta_A) = 0.7$ | |

In an evidential database, as shown in Table 1, an item corresponds to a focal element. An itemset corresponds to a conjunction of focal elements having different domains. Two different itemsets can be related via the inclusion or the intersection operator. Indeed, the inclusion operator for evidential itemsets [9] is defined as follows, let $X$ and $Y$ be two evidential itemsets:

$$X \subseteq Y \iff \forall x_i \in X, x_i \subseteq y_i.$$

where $x_i$ and $y_i$ are the $i^{th}$ element of $X$ and $Y$. For the same evidential itemsets $X$ and $Y$, the intersection operator is defined as follows:

$$X \cap Y = Z \iff \forall z_i \in Z, z_i \subseteq x_i \text{ and } z_i \subseteq y_i.$$

An *Evidential associative rule* $R$ is a causal relationship between two itemsets that can be written in the following form $R : X \rightarrow Y$ fulfilling $X \cap Y = \emptyset$. In Table 1, $A_1$ is an item and $\{\theta_A\ B_1\}$ is an itemset such that $A_1 \subset \{\theta_A\ B_1\}$ and $A_1 \cap \{\theta_A\ B_1\} = A_1$. $A_1 \rightarrow B_1$ is an evidential associative rule.

Several definitions for the support estimation were defined for the evidential itemsets such as [8,9]. Those methods assess the support based on the belief function applied on the evidential database BBA $m_{\mathcal{EDB}}$[1] :

$$Support_{\mathcal{EDB}}(X) = Bel_{\mathcal{EDB}}(X) \tag{2}$$

such that:

$$Bel : 2^{\theta} \rightarrow [0, 1] \tag{3}$$

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B). \tag{4}$$

In a previous work [10], we introduced a new metric for support estimation providing more accuracy and overcoming several limits of using the belief function. The Precise support $Pr$ is defined by:

$$Pr : 2_i^{\theta} \rightarrow [0, 1] \tag{5}$$

$$Pr(x_i) = \sum_{x \subseteq \theta_i} \frac{|x_i \cap x|}{|x|} \times m_{ij}(x) \qquad \forall x_i \in 2^{\theta_i}. \tag{6}$$

The evidential support of an itemset $X = \prod_{i \in [1 \ldots n]} x_i$ in the transaction $T_j$ (i.e., $Pr_{T_j}$) is then computed as follows:

$$Pr_{T_j}(X) = \prod_{x_i \in \theta_i, i \in [1 \ldots n]} Pr(x_i) \tag{7}$$

[1] A BBA constructed from Cartesian product applied on the evidential database. Interested readers may refer to [8].

Thus, the evidential support $Support_{\mathcal{EDB}}$ of the itemset $X$ becomes:

$$Support_{\mathcal{EDB}}(X) = \frac{1}{d}\sum_{j=1}^{d} Pr_{T_j}(X). \tag{8}$$

## 3 Confidence Measure for Evidential Associative Rules

The confidence is the measure assigned to the associative rules and it represents its relevance [1]. As originally introduced in Boolean databases, the confidence measure was relying on conditional probability [1]. Indeed for a rule $R : R_a \rightarrow R_c$, such that $R_c$ and $R_a$ are respectively the conclusion and the antecedent (premise) part of the rule $R$, the confidence is expressed as follows:

$$Confidence(R) = P(R_c|R_a) = \frac{\sum_{i=1}^{d} P(R_a \cap R_c)}{\sum_{i=1}^{d} P(R_a)} \tag{9}$$

In addition, even in fuzzy data mining, the associative rule's confidence is built with conditional fuzzy measures [11]. In this respect, evidential associative rules were initially introduced in [8]. The authors defined the structure of an evidential associative rule and estimated its relevance following a confidence metric. The confidence of a rule $R$ in the set of all rules $\mathcal{R}$, i.e., $R \in \mathcal{R}$, is computed as follows:

$$Confidence(R) = Bel(R_c|R_a) \tag{10}$$

where $Bel(\bullet|\bullet)$ is the conditional Belief. The proposed confidence metric is hard to define where several works have tackled this issue and different interpretations and formulas were proposed such as those given respectively in [5,12]. In [5], the conditional belief is defined as follows:

$$Bel(R_c|R_a) = \frac{Bel(R_c \cup \overline{R_a}) - Bel(\overline{R_a})}{1 - Bel(\overline{R_a})} \tag{11}$$

In [8], the authors used Fagin et al.'s conditional belief such that:

$$Bel(R_c|R_a) = \frac{Bel(R_a \cap R_c)}{Bel(R_a \cap R_c) + Pl(R_a \cap \bar{R}_c)}. \tag{12}$$

where $Pl()$ is the plausibility function and is defined as follows:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \tag{13}$$

*Example 1.* Through the following example, we highlight the inadequacy of the conditional belief use. We consider the Transaction 1 of Table 1 from which we

try to compute the confidence of $A_2 \rightarrow B_1$ (i.e., $Bel(B_1|A_2)$). The conditional belief introduced in [5] gives the following results:

$$Bel(B_1|A_2) = \frac{Bel(B_1 \cup \overline{A_2}) - Bel(\overline{A_2})}{1 - Bel(\overline{A_2})} = \frac{Bel(B_1)}{1} = 0.4$$

The result of the belief of $B_1$ knowing $A_2$ is true is equal to that of $Bel(B_1)$ due to the independence between $A_2$ and $B_1$. On the other hand, both hypothesis might be correlated so that the event $B_1$ does not occur knowing already the happening of $A_2$.

In the following, we propose a new metric for the confidence estimation based on our Precise support measure [10] and probability assumption:

$$Confidence(R) = \frac{\sum_{j=1}^{d} Pr_{T_j}(R_a) \times Pr_{T_j}(R_c)}{\sum_{j=1}^{d} Pr_{T_j}(R_a)} \tag{14}$$

where $d$ is the number of transactions in the evidential database. Thanks to its probabilistic writing, the proposed metric sustains previous confidence measure such as that introduced in [1].

*Example 2.* Let us consider the example of the evidential database in Table 1. The confidence of the evidential associative rule $R_1 : A_1 \rightarrow B_1$ is computed as follows:

$$Confidence(R_1) = \frac{Pr_{T_1}(A_1) \times Pr_{T_1}(B_1) + Pr_{T_2}(A_1) \times Pr_{T_2}(B_1)}{Pr_{T_1}(A_1) + Pr_{T_2}(A_1)} = 0.75$$

The generated rules with their confidence could find several applications. In the following, we tackle the classification problem case and a based evidential rule classifier is introduced.

## 4 Associative Rule Classifier

One of the main characteristics of the evidential database is the great number of items that it integrates. The number of items depends from the frame of discernment of each column. This asset makes from the evidential database more informative but more complex than the usual binary database. In [10], we have shown the significant number of generated frequent patterns that may be drawn even from small databases. Indeed, from a frequent itemset, of size $k$, $2^k - 2$ potential rules are generated. In order to use the generated evidential rules for a classification purposes, we first have to reduce their number for a more realistic one. In the following, we propose two processes for classification rule's reduction.

### 4.1 Classification Rules

From the obtained rules, we retain only the classification ones. From a rule such that $\prod_{i \in I} X_i \rightarrow \prod_{j \in J} Y_j$, we only keep those matching a class hypothesis at the conclusion part (i.e., $Y_j \in \theta_C$ and $\theta_C$ is the frame of discernment).

*Example 3.* Let us consider the following set of the association rules $S = \{A_1 \rightarrow C_1; A_1, B_2 \rightarrow C_1; A_1 \rightarrow B_1\}$ and the class frame of discernment $\theta_C = \{C_1, C_2\}$. After classification rule reduction, the set $S$ becomes $S = \{A_1 \rightarrow C_1; A_1, B_2 \rightarrow C_1\}$.

### 4.2 Generic and Precise Rules

**Generic Rules:** the rule's reduction can assimilate the redundant rules. A rule $R_1$ is considered as a redundant rule if and only if it does not bring any new information having at hand a rule $R_2$. $R_2$ is considered as more informative as far as its antecedent part is included in that of $R_1$. The retained rules from the reduction process constitute the set of Generic rules $\mathcal{R}$ extracted from the set of frequent itemsets $\mathcal{FI}$.

*Example 4.* Let us consider the previous set of the association rules $S = \{A_1 \rightarrow C_1; A_1, B_2 \rightarrow C_1; A_1 \rightarrow B_1\}$. After redundant rule reduction, the set $S$ becomes $S = \{A_1 \rightarrow C_1; A_1 \rightarrow B_1\}$.

**Precise Rules:** A rule is considered as *precise* if the rule's premise is maximized. Thus, from the set of all possible rules, we retain only those having the size of their premise part equal to $n$ (number of columns of $\mathcal{EDB}$).

Algorithm 1 sketches the process of rule's generation as well as rule reduction. The algorithm relies on the function $Construct\_Rule(x, \theta_C)$ (Line 10) that generates associative rules and filtrates out them by retaining only the classification ones. The function $Find\_Confidence(R, Pr\_Table)$ (Line 22) computes the confidence of the rule $R$ following the $Pr\_Table$ that contains all transactional support of each item (for more details see [10]). Finally, the function $Redundancy(\mathcal{R}, R)$ (Line 42) builds the set of all classification rules $\mathcal{R}$ which are not redundant and having the confidence value greater than or equal to the fixed threshold $minconf$.

### 4.3 Classification

Let us suppose the existence of an instance $X$ to classify represented a set of BBA belonging to the evidential database $\mathcal{EDB}$ such that:

$$X = \{m_i | m_i \in X, x_i^j \in \theta_i\} \tag{15}$$

where $x_i^j$ is a focal element of the BBA $m_i$. Each retained associative rule, in the set of rules $\mathcal{R}$, is considered as a potential piece of information that could be of

**Algorithm 1.** Evidential Associative Rule Generation algorithm

**Require:** $Pr\_Table, minconf, \mathcal{FI}, \theta_C$
**Ensure:** $\mathcal{R}$
**for all** $x \in \mathcal{FI}$ **do**
  $R \leftarrow Construct\_Rule(x, \theta_C)$
  **if** $R \neq \emptyset$ **then**
    $Conf \leftarrow Find\_Confidence(R, Pr\_Table)$
    **if** $Conf > minconf$ **then**
      $\mathcal{R} \leftarrow Redundancy(\mathcal{R}, R)$
    **end if**
  **end if**
**end for**
**function** CONSTRUCT_RULE($X$ , $\theta_C$)
  **for all** $x \in X$ **do**
    **if** $x \notin \theta_C$ **then**
      $prem \leftarrow prem + \{x\}$
    **else**
      $concl \leftarrow concl + \{x\}$
    **end if**
  **end for**
  $R.premise \leftarrow prem$
  $R.conclusion \leftarrow concl$
  **return** $R$
**end function**
**function** FIND_CONFIDENCE($R$, $Pr$)
  $numer \leftarrow 0$
  $denom \leftarrow 0$
  **for** j=1 to d **do**
    $num \leftarrow 1$
    $den \leftarrow 1$
    **for all** $i \in Pr(j).focal\_element$ **do**
      **if** $Pr(j).focal\_element \in R.premise$ **then**
        $num \leftarrow num \times Pr(j).val$
        $den \leftarrow den \times Pr(j).val$
      **else**
        **if** $Pr(j).focal\_element \in R.conclusion$ **then**
        **end if**
      **end if**
    **end for**
    $numer \leftarrow numer + num$
    $denom \leftarrow denom + den$
  **end for**
  **return** $\frac{numer}{denom}$
**end function**
**function** REDUNDANCY($\mathcal{R}$,$R$)
  **for all** $rule \in \mathcal{R}$ **do**
    **if** $R.premise \subset rule.premise$ & $R.conclusion = rule.conclusion$ **then**
      $\mathcal{R} \leftarrow \mathcal{R} \backslash rule$
      $\mathcal{R} \leftarrow \mathcal{R} \cup R$
    **end if**
  **end for**
  **return** $\mathcal{R}$
**end function**

help for $X$ class determination. In order to select rules that may contribute to classification, we look for rules having a non null intersection with $X$ such that:

$$\mathcal{RI} = \{R \in \mathcal{R}, \exists x_i^j \in \theta_i,\ x_i^j \in R_a\} \tag{16}$$

Each rule found in the set $\mathcal{RI}$ constitutes a piece of information concerning the instance $X$ membership. Since several rules can be found and fulfilling the intersection condition, it is of importance to benefit from them all. In our work, we assume that all information is valuable and should be handled within the information fusion problem. From the set $\mathcal{RI}$, we extract the set of generic or precise classification rules (see Subsection 4.2). Indeed, each rule from the computed set $R^l \subset \mathcal{RI}, l \in [1 \ldots L]$ and $L < |\mathcal{RI}|$, that brings a new information

(different $R_a$) is transformed into a BBA following the frame of discernment $\theta_C$ (frame of discernment of $R_c$):

$$\begin{cases} m_{R^l}^{\theta_C}(\{R_c\}) = \alpha \times confidence(R^l) \\ m_{R^l}^{\theta_C}(\theta_C) = 1 - (\alpha \times confidence(R^l)) \end{cases} \tag{17}$$

where $R_c$ is the conclusion part of the rule $R^l$ and $\alpha \in [0, 1]$ is a discounting factor.

The $L$ constructed BBA are then fused following the Dempster rule of combination [5] as follows:

$$m_{\oplus} = \oplus_{l=1}^{L} m_{R^l}^{\theta_C}. \tag{18}$$

$\oplus$ is the Dempster's aggregation function where for two source's BBA $m_1$ and $m_2$:

$$\begin{cases} m_{\oplus}(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) \cdot m_2(C) & \forall A \subseteq \Theta, A \neq \emptyset \\ m_{\oplus}(\emptyset) = 0 \end{cases} \tag{19}$$

where $K$ is defined as:

$$K = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C). \tag{20}$$

## 5 Experimentation and Results

In this section, we present how we managed to conduct our experiments and we discuss comparative results.

### 5.1 Evidential Database Construction

In order to perform experimental tests, we construct our own evidential databases from UCI benchmarks [13] based upon ECM [14]. Interested reader may refer to [10] for more details on evidential database construction. The transformation was operated on *Iris*, *Vertebral Column*, *Diabetes* and *Wine* databases. The studied databases are summarized on Table 2 in terms of number of instances and attributes.

**Table 2.** Database characteristics

| **Database** | **#Instances** | **#Attributes** | **#Focal elements** |
|---|---|---|---|
| Iris_EDB | 150 | 5 | 40 |
| Vertebral Column_EDB | 310 | 7 | 116 |
| Diabetes_EDB | 767 | 9 | 132 |
| Wine_EDB | 178 | 14 | 196 |

### 5.2 Comparative Results

In the following, we compare the classification result performance between the Generic and Precise rules. Table 3 shows the difference in classification result between the generic and the precise associative rules. The precise rules highlight better results than do the generic ones. Indeed, the larger the rule's premise is, the more pertinent the rule is. On the other hand, the generic rule based approach fuse much more rules than do the precise one. In addition, all generic rules are considered with the same weight within the fusion process despite their pertinence difference. These characteristics with Dempster's combination behavior mislead the fusion process to errors. Indeed, as shown in Figure 1, the high number of fused rules depends highly from the *minsup* value. Unlike the generic approach, the number of precise rule is defined by number of larger premise's rule which is dependent from the treated evidential transaction.

**Table 3.** Comparative result between Generic and Precise classification rules

| Database | Iris_EDB | Vertebral Column_EDB | Diabetes_EDB | Wine_EDB |
|---|---|---|---|---|
| Precise rules | 80.67% | 88.38% | 83.20% | 100% |
| Generic rules | 78.67% | 67.74% | 65.10% | 51.68% |

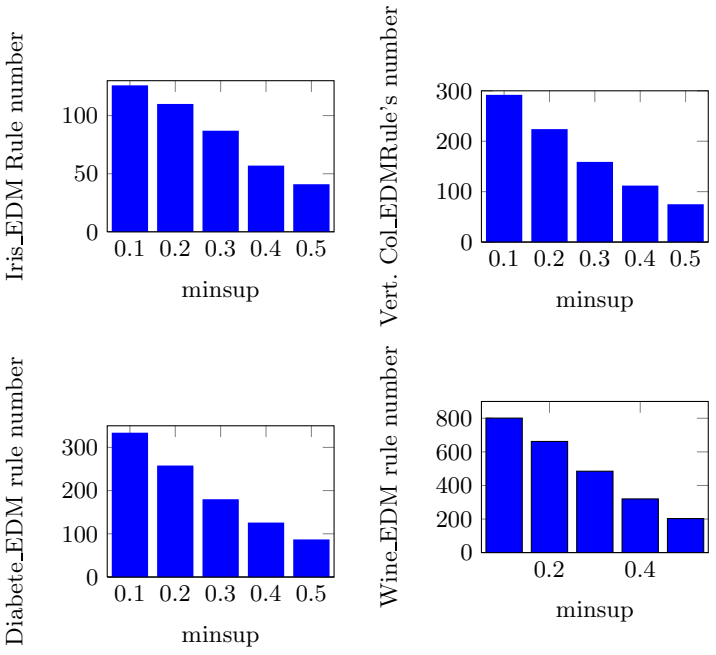



**Fig. 1.** Generic associative rule's number for different support values

## 6 Conclusion

In this paper, we tackled associative rule's extraction from evidential databases. We proposed a new confidence measure for associative rules in evidential databases. The proposed measure is based on Precise support (i.e., probability measure) providing coherence and sustains previous work on fuzzy and binary databases. The rules are then filtrated to retain only classification and non redundant rules. A classification method based on evidential associative rules is introduced. The classification approach is based on a fusion system that represents interesting rules. As illustrated in the experimentation section, the proposed method provides an interesting performance rates. In future work, we plan to study the development of a new method to estimate the reliability of each combined associative rule. Indeed, each rule has a precision relatively to the instance to classify. The precision is measured by the intersection between the premise and the instance itself. A reliability measure for rule BBA is under study.

## References

1. Agrawal, R., Srikant, R.: Fast algorithm for mining association rules. In: Proceedings of International Conference on Very Large DataBases, VLDB, Santiago de Chile, Chile, pp. 487–499 (1994)
2. Li, W., Han, J., Pei, J.: CMAR: Accurate and efficient classification based on multiple class-association rules. In: Proceedings of IEEE International Conference on Data Mining (ICDM 2001), pp. 369–376. IEEE Computer Society, San Jose (2001)
3. Bouzouita, I., Elloumi, S., Ben Yahia, S.: GARC: A new associative classification approach. In: Proceedings of 8th International Conference on Data Warehousing and Knowledge Discovery, DaWaK, Krakow, Poland, pp. 554–565 (2006)
4. Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)
5. Dempster, A.: Upper and lower probabilities induced by multivalued mapping. AMS-38 (1967)
6. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
7. Lee, S.: Imprecise and uncertain information in databases: an evidential approach. In: Proceedings of Eighth International Conference on Data Engineering, Tempe, AZ, pp. 614–621 (1992)
8. Hewawasam, K.K.R., Premaratne, K., Shyu, M.-L.: Rule mining and classification in a situation assessment application: A belief-theoretic approach for handling data imperfections. Trans. Sys. Man Cyber. Part B 37(6), 1446–1459 (2007)
9. Bach Tobji, M.A., Ben Yaghlane, B., Mellouli, K.: Incremental maintenance of frequent itemsets in evidential databases. In: Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Verona, Italy, pp. 457–468 (2009)
10. Samet, A., Lefevre, E., Ben Yahia, S.: Mining frequent itemsets in evidential database. In: Proceedings of the Fifth International Conference on Knowledge and Systems Engeneering, Hanoi, Vietnam, pp. 377–388 (2013)

11. Hong, T.-P., Kuo, C.-S., Wang, S.-L.: A fuzzy AprioriTid mining algorithm with reduced computational time. Applied Soft Computing 5(1), 1–10 (2004)
12. Fagin, R., Halpern, J.Y.: A new approach to updating beliefs. Uncertainty in Artificial Intelligence, 347–374 (1991)
13. Frank, A., Asuncion, A.: UCI machine learning repository (2010), http://archive.ics.uci.edu/ml
14. Masson, M.-H., Denœux, T.: ECM: An evidential version of the fuzzy c-means algorithm. Pattern Recognition 41(4), 1384–1397 (2008)

# Towards Evidence-Based Terminological Decision Trees

Giuseppe Rizzo, Claudia d'Amato, Nicola Fanizzi, and Floriana Esposito

LACAM – Università degli studi di Bari "Aldo Moro", Bari, Italy
name.lastname@uniba.it

**Abstract.** We propose a method that combines terminological decision trees and the Dempster-Shafer Theory, to support tasks like *ontology completion*. The goal is to build a predictive model that can cope with the epistemological uncertainty due to the Open World Assumption when reasoning with Web ontologies. With such models not only one can predict new (non derivable) assertions for completing the ontology but by assessing the quality of the induced axioms.

## 1 Introduction

In the context of machine learning applied to Web ontologies, various methods have been proposed in order to predict new assertions. It has turned out that the models resulting from these methods can often provide effective reasoning services which are comparable to those offered by reasoners [1].

Focusing on the *instance-checking* problem [2], i.e. the task of assessing the class-membership for individual resources, it is well known that a reasoner may be unable to prove the membership of an individual to a given concept (or to its complement). This is often caused by flaws introduced in the ontology construction phase owing to lacking disjointness axioms. The same problem may appear also with logic-based predictive classification models produced by machine learning algorithms, such as the *Terminological Decision Trees* [3] (TDTs), a specialization of first-order decision trees [4]. In this work we extend the scope of TDTs by employing the *Dempster-Shafer Theory* [5] (DST) because, differently from the instance-based approach proposed in [6], logic-based classification models generally do not provide an epistemic uncertainty measure. This may be very important as a quality measure for predicted assertions in related problems such as data integration in the context of Linked Data[1], where it could contribute as a measure of *provenance* [7]. Purely logical models cannot handle properly cases of tests resulting in an unknown membership. The uncertainty is not explicitly considered when an individual is classified w.r.t. a given test class. The situation is similar to the case of missing values in prediction with (propositional) decision trees. The underlying idea of the proposed extension is to exploit standard algorithms to cope with missing values for a test by partitioning the observation w.r.t. all possible values of the test and then following all branches. Once the leaves are reached, the results are combined. Thanks to the combination rules used to pool evidences [5], the DST is a more suitable framework than the Bayesian theory of probability to cope with epistemic uncertainty

[1] http://linkeddata.org

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 36–45, 2014.

and ignorance related to the Open World Assumption (OWA) that characterizes Web ontologies.

The DST has been integrated in various algorithms [8,9] with results that are competitive with the classical version. So, we want to investigate if this model can be used also in machine learning algorithm for the Semantic Web in order to obtain better results of classifiers in terms of predicted assertions.

The paper is organized as follows: Section 2 introduces basics concerning the concept learning task in *Description Logic knowledge bases* and describes the original version of TDTs; in Section 3 the algorithm for inducing a TDT based on the DST is proposed while in Section 4 an early-stage empirical evaluation is described; finally, further extensions of this work are described.

## 2 Background

**Knowledge Bases.** In Description Logics (DLs) [2], a domain is modeled through primitive *concepts* (classes) and *roles* (relations), which can be used to build complex descriptions regarding *individuals* (instances, objects), by using specific operators that depend on the adopted language. A *knowledge base* is a couple $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ where the *TBox* $\mathcal{T}$ contains axioms concerning concepts and roles (typically inclusion axioms such as $C \sqsubseteq D$) and the *ABox* $\mathcal{A}$ contains assertions, i.e. axioms regarding the individuals ($C(a)$, resp. $R(a, b)$). The set of individuals occurring in $\mathcal{A}$ is denoted by $\mathsf{Ind}(\mathcal{A})$.

The semantics of concepts/roles/individuals is defined through interpretations. An *interpretation* is a couple $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ where $\Delta^{\mathcal{I}}$ is the *domain* of the interpretation and $\cdot^{\mathcal{I}}$ is a *mapping* such that, for each individual $a$, $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, for each concept $C$, $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and for each role $R$, $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The semantics of complex descriptions descends from the interpretation of the primitive concepts/roles and of the operators employed, depending on the adopted language. $\mathcal{I}$ satisfies an axiom $C \sqsubseteq D$ ($C$ is subsumed by $D$) when $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ and an assertion $C(a)$ (resp. $R(a, b)$) when $a^{\mathcal{I}} \in C^{\mathcal{I}}$ (resp. $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$). $\mathcal{I}$ is a *model* for $\mathcal{K}$ iff it satisfies each axiom/assertion $\alpha$ in $\mathcal{K}$, denoted with $\mathcal{I} \models \alpha$. When $\alpha$ is satisfied w.r.t. these models, we write $\mathcal{K} \models \alpha$.

We will be interested in the *instance-checking* inference service: given an individual $a$ and a concept description $C$ determine if $\mathcal{K} \models C(a)$. Due to the *Open World Assumption* (OWA), answering to a class-membership query is more difficult w.r.t. *Inductive Logic Programming* (ILP) settings where the closed-world reasoning is the standard. Indeed, one may not be able to prove the truth of either $\mathcal{K} \models C(a)$ or $\mathcal{K} \models \neg C(a)$, as there may be possible to find different interpretations that satisfy either cases.

**Learning Concepts in DL.** The concept learning task in DL can be defined as follows. Given:

- a knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$
- a target concept $C$,
- the sets of positive and negative examples for $C$:
  $Ps = \{a \in \mathsf{Ind}(\mathcal{A}) \mid \mathcal{K} \models C(a)\}$ and $Ns = \{a \in \mathsf{Ind}(\mathcal{A}) \mid \mathcal{K} \models \neg C(a)\}$

the goal is to obtain a concept description $D$ for $C$ ($C \sqsubseteq D$), such that:

- $\mathcal{K} \models D(a) \quad \forall a \in Ps$
- $\mathcal{K} \models \neg D(a) \quad \forall a \in Ns$

In order to investigate about the learning of multiple disjoint concepts, the formulation of the problem is more restrictive than the one proposed in [3], where the negative examples were the individuals $a$ for which $\mathcal{K} \not\models D(a)$ . The resulting concept descriptions can be used to solve an *instance-checking* problem for new individuals. Similarly to a *First Order Logic Decision Tree*, a binary tree where each node contains a conjunction of literals and each variable that is introduced in a node cannot appear in the right branch of that node, a Terminological Decision Tree can be defined as follows:

**Definition 1 (Terminological Decision Tree).** *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, a* Terminological Decision Tree *is a binary tree where:*

- *each node contains a conjunctive concept description $D$;*
- *each departing edge is the result of a class-membership test w.r.t. $D$, i.e., given an individual $a$, $\mathcal{K} \models D(a)$?*
- *if a node with $E$ is the father of the node with $D$ then $D$ is obtained by using a refinement operator and one of the following conditions should be verified:*
  - *$D$ introduces a new concept name,*
  - *$D$ is an existential restriction,*
  - *$D$ is an universal restriction of any its ancestor.*

However, a set of concept descriptions is generated by means of the refinement operator and the best one is chosen to be installed as a child node. The best description is the one that maximizes a purity measure respect to the previous level [3]. The measure may be defined as *accuracy* in a binary classification problem, *purity* $= p/(p+n)$, where $p$ is the number of positive examples and $n$ the number of negative ones reaching a node.

## 3 Induction of the Terminological Trees

The method for inducing TDTs based on the DST uses a divide-and-conquer strategy. It requires the target concept $C$, a training set $Ps \cup Ns \cup Us$ made up of individuals with positive ($Ps$), negative($Ns$) and uncertain ($Us$) membership w.r.t. $C$ and a *basic belief assignment* (BBA) $m$ associated with $C$ (with $\Omega = \{+1, -1\}$ as frame of discernment).

The main learning function (see Alg. 1) refines the input test concept using one of the available operators. After candidates are generated, a BBA is computed for each of them. The BBAs for the node concepts are simply estimated based on the number of positive, negative and uncertain instances in the training set:

- $m(\{+1\}) \leftarrow |Ps|/|Ps \cup Ns \cup Us|$;
- $m(\{-1\}) \leftarrow |Ns|/|Ps \cup Ns \cup Us|$;
- $m(\Omega) \leftarrow |Us|/|Ps \cup Ns \cup Us|$;

*Example 2 (Computation of a BBA).* Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, consider the concepts `Man`, and its complement $\texttt{Woman} \equiv \neg\texttt{Man}$ in $\mathcal{T}$ and the following assertions:
$\{\texttt{Man(BOB)}, \texttt{Man(JOHN)}, \texttt{Woman(ANN)}, \texttt{Woman(MARY)}\} \subset \mathcal{A}$, with `MARK` occurring elsewhere in $\mathcal{A}$ and whose membership w.r.t. `Man` is unknown. A BBA for `Man` is produced (the frame of discernment $\Omega_{\texttt{Man}} = \{+1, -1\}$ corresponds to $\{\texttt{Man}, \texttt{Woman}\}$):

**Algorithm 1.** Induction of DST-based TDT

```
1  input: Ps, Ns, Us: set of training instances {positive, negative, uncertain membership}
2         C: concept description,
3         m: BBA
4  output: T: TDT
5  const θ, ν ∈ [0, 1] {thresholds}
6  function INDUCEDSTTDTREE(Ps, Ns, Us:individuals; C:concept; m: BBA)
7  begin
8      if |Ps| = 0 and |Ns| = 0 then
9          begin
10         if Pr+ ≥ Pr− then {pre−defined constants wrt the whole training set}
11             T_root ← ⟨C, m⟩
12         else
13             T_root ← ⟨¬C, m⟩
14         return T
15         end
16     if m({−1} = 0 and m({+1}) > θ then
17         begin
18         T_root ← ⟨C, m⟩
19         return T
20         end
21     if m({+1} = 0 and m({−1}) > θ then
22         begin
23         T_root ← ⟨¬C, m⟩
24         return T
25         end
26     if NONSPECIFITY(C) ≥ ν then
27         begin
28         if m({+1}) ≥ m({−1}) then
29             T_root ← ⟨C, m⟩
30         else
31             T_root ← ⟨¬C, m⟩;
32         return T
33         end
34     S ← GENERATECANDIDATES(Ps, Ns, Us)
35     D ← SELECTBESTCONCEPT(m, S)
36     ⟨⟨Pl, Nl, Ul⟩, ⟨Pr, Nr, Ur⟩⟩ ← SPLIT(D, Ps, Ns, Us)
37     T_root ← ⟨D, m_D⟩
38     T_left ← INDUCEDSTTDTREE(Pl, Nl, Ul, D, m_D)
39     T_right ← INDUCEDSTTDTREE(Pr, Nr, Ur, D, m_D)
40     return T
41 end
```

- $m(\{+1\}) = |\{\texttt{BOB}, \texttt{JOHN}\}|/|\{\texttt{BOB}, \texttt{JOHN}, \texttt{ANN}, \texttt{MARY}, \texttt{MARK}\}| = \frac{2}{5} = 0.4$
- $m(\{-1\}) = |\{\texttt{ANN}, \texttt{MARY}\}|/|\{\texttt{BOB}, \texttt{JOHN}, \texttt{ANN}, \texttt{MARY}, \texttt{MARK}\}| = \frac{2}{5} = 0.4$
- $m(\Omega_{\texttt{Man}}) = |\{\texttt{MARK}\}|/|\{\texttt{BOB}, \texttt{JOHN}, \texttt{ANN}, \texttt{MARY}, \texttt{MARK}\}| = \frac{1}{5} = 0.2$ □

The set of candidates $S$, is made up of pairs $\langle D, m_D\rangle$ where $D$ is a concept description and $m_D$ is the BBA computed for it. After $S$ has been generated, the algorithm selects the best test concept and the corresponding BBA according to measure computed from the BBA and the best pair $\langle D, m_D\rangle$ is installed as a child node of $\langle C, m\rangle$.

This strategy is repeated recursively, splitting the examples according to the test concept in each node. Recursion stops when only positive (resp. negative) instances are rooted to a node which becomes a leaf (see the conditions checked in lines 16 and 21 in Alg. 1). The first condition (line 8) refers to the case when no positive and negative instances reach the node. In this case the algorithm uses priors, $Pr^+$ and $Pr^-$, precomputed for the whole training set. The fourth case expresses a situation in which the child nodes added to the tree are characterized by a high non-specificity measure.

**Algorithm 2.** Candidate test concepts generation

**input**: $Ps$, $Ns$, $Us$: set of training instances
  $C$: concept description
**output**: $S$: set of $\langle D, m_D \rangle$ {$D$ *is a concept and* $m$ *is a BBA*}
**function** GENERATECANDIDATES ($Ps$, $Ns$, $Us$, $C$): $S$
**begin**
$S \leftarrow \emptyset$
$L \leftarrow$ GENERATEREFINEMENTS ($C$, $Ps$, $Ns$) {*based on the refinement operator*}
**for each** $D \in L$
  **begin**
  $m \leftarrow$ GETBBA ($Ps_D$, $Ns_D$, $Us_D$)
  $S \leftarrow S \cup \{\langle D, m \rangle\}$
  **end**
**return** $S$
**end**

**Algorithm 3.** Selection of the best candidate

**input**: $m_C$: BBA,
  $S$: set of $\langle D, m_D \rangle$
**output**: $C'$: concept {*best according to non−specificity measure*}
**const** $\Omega$, $\Omega_D$ ($\forall D \in S$): frame of discernment
**function** SELECTBESTCANDIDATE ($m_C$, $S$): concept
**begin**
  $Ns \leftarrow \sum_{A \subseteq \Omega} m_C(A) \log |A|$
  $C' \leftarrow \text{argmax}_{D \in S} \left( Ns - \sum_{A' \subseteq \Omega_D} m_D(A') \log |A'| \right)$
  **return** $C'$
**end**

For a given concept description $D$, NONSPECIFICITY(D) is the value computed from its BBA $\sum_{A \in 2^\Omega} m(A) \log(|A|)$ as a measure of imprecision [9]. The algorithm controls the growth by means of the threshold $\nu$. If the condition is verified, the algorithm compares $m(\{+1\})$ and $m(\{-1\})$ to install the proper test concept in the node.

Alg. 2 illustrates how a set $S$ of candidate concepts is generated by GENERATECANDIDATES. This function calls GENERATEREFINEMENTS to generate refinements that can be used as tests. $S$ is updated with pairs $\langle D, m_D \rangle$ where $D$ is a refinement and $m_D$ is the related BBA. Once concepts have been generated, SELECTBESTCANDIDATE (see Alg. 3) selects the best candidate description according to the *non-specificity* measure. The advantage of this method is the explicit representation of the OWA using the maximal ignorance hypothesis (i.e. the one corresponding to $\Omega$).

**BBA Creation.** As previously mentioned, the proposed approach associates a BBA to each node of a TDT for representing the epistemic uncertainty about the class-membership. The BBA of the child node is created from a subset of the training examples routed to the parent node.

When a branch is created together with the related concept description, the membership of the individuals w.r.t. this concept is computed in order to obtain a BBA whose frame of discernment represents the hypothesis of membership w.r.t. that concept.

Moreover, when a new node is added as left or right child, the algorithm knows about the tests performed on the parent node concept for each instance. Hence, similarly to the Bayesian framework, an implicit kind of conditioning results that allows to relate

the membership w.r.t. a concept description in a parent node to the membership w.r.t. the refinements contained in its children.

**Stop Criterion.** We can apply the DST to TDTs to decide when to stop the growth. As described above, we add a new child node minimizing the degree of imprecision represented by the non-specificity measure. However, if the number of instances with unknown-membership is very high for the new refinement, the imprecision increases and the node should not be further refined. Thus, the algorithm uses two stop criteria:

**purity** $m(\{-1\}) = 0 \wedge m(\{+1\}) > \theta$ or $m(\{+1\}) = 0 \wedge m(\{-1\}) > \theta$
**non-specificity:** given a description $D$, $\text{NONSPECIFICITY}(D) > \nu$

The former condition derives from decision tree induction, where a leaf is formed when only instances that belong to a single class remain. In terms of DST, this idea can be represented by a BBA where:

$$\forall A \in 2^{\Omega} \quad m(A) = \begin{cases} 1 & \text{if } A = \{+1\} \text{ (\textit{resp.} } A = \{-1\}) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Thus, the first condition distinguishes two kinds of individuals: those with a certain membership (positive, negative) and those with an uncertain membership. The latter condition moves from the idea that once the best candidate concept description has been chosen, it can be very imprecise (i.e. the measure of non-specificity is very high). Thus, the resulting BBA has the maximum value assigned to a the case of total ignorance w.r.t. the BBA of the parent concept. The threshold $\nu$ is used to control this condition.

**Prediction.** After the TDT has been produced, it can be used to predict the class-membership in the usual way. Given an individual $a \in \mathsf{Ind}(A)$, a path is followed down the tree according to the results of the test w.r.t. the concept $D$ at each node.

Alg. 4 describes the recursive strategy. The base case is when a leaf is reached. In this case, the algorithm updates a list with the BBA associated with the leaf node. The recursive step follows a branch rather than another according to the result of instance-checking w.r.t. the concept description $D$. If $\mathcal{K} \models D(a)$ the algorithm follows the left branch of the tree. If $\mathcal{K} \models \neg D(a)$ the right branch is followed. A more interesting case occurs when the result of instance check is unknown, i.e. $\mathcal{K} \not\models D(a)$ or $\mathcal{K} \not\models \neg D(a)$. In this case, both the left and the right branch are followed until the leaves are reached. In this way the algorithm can cope with the OWA. The underlying idea is to collect all the possible classifications when the result of a test on an internal node is unknown. In these cases the DST seems to be a good framework in order to combine all such results and make a decision on the membership to be assigned.

After the tree exploration, we may have various BBAs in the list (one per reached leaf). Then, these functions are to be pooled according to a *combination rule* (see Alg. 5). The resulting BBA can be used to compute *belief*, *plausibility* or *confirmation* [5] on the membership hypothesis and the algorithm returns the hypothesis that maximize one of them. Similarly to our previous works [6], we considered the computation of the confirmation in order to balance belief and plausibility in the final decision.

**Algorithm 4.** Determining the class-membership of an individual

```
input: a: test individual,
       T: TDT,
       K: knowledge base
output: L: list of BBA {related to the leave nodes}
function FINDLEAVES(a, T, K) : L
begin
    N ← ROOT(T)
    if ¬LEAF(N, T) then
        begin
        ⟨D, T_left, T_right⟩ ← INODE(N);
        if K ⊨ D(a) then
            L ← FINDLEAVES(a, T_left, K)
        else if K ⊨ ¬D(a) then
            L ← FINDLEAVES(a, T_right, K)
        else
            begin
            L ← FINDLEAVES(a, T_left, K)
            L ← FINDLEAVES(a, T_right, K)
            end
        end
    else
        begin
        m ← GETBBA(N)
        L ← L ∪ {m}
        end
    return L
end
```

# 4 Preliminary Experiments

## 4.1 Setup

The goal of the experimentation is to evaluate the TDTs induced by the proposed method in the class-membership prediction task. We considered various Web ontologies (see Tab. 1). For each of them, 30 query concepts have been randomly generated by combining (using the conjunction and disjunction operators or universal and existential restriction) 2 through 8 concepts of the ontology. A *0.632 bootstrap* procedure was employed for the design of the experiments.

The experiments were repeated under three different conditions. First we ran the original method for learning TDTs. Then, we ran them with the DST-based version

**Algorithm 5.** Pooling evidence for classification

```
input: a: individual,
       T: TDT,
       K: knowledge base
output: v ⊆ {−1, +1}
function CLASSIFY(a, T, K): v
begin
    L ← FINDLEAVES(a, T, K) {list of BBA}
    m̄ ← ⊕_{m∈L}
    for v ∈ Ω do
        compute Bel̄_v and Pl̄_v
    Conf̄_v ← CONFIRMATION(Bel̄_v, Pl̄_v)
    return argmax_{v⊆Ω} Conf̄_v
end
```

**Table 1.** Ontologies employed in the experiments

| *Ontology* | *Expressiivity* | *Concepts* | *Roles* | *Individuals* |
|---|---|---|---|---|
| FSM | $\mathcal{SF}(D)$ | 20 | 10 | 37 |
| LEO | $\mathcal{ALCHIF}(D)$ | 32 | 26 | 62 |
| LUBM | $\mathcal{ALEHIF}(D)$ | 43 | 25 | 1555 |
| BIOPAX | $\mathcal{ALCIF}(D)$ | 74 | 70 | 323 |
| NTN | $\mathcal{SHIF}(D)$ | 47 | 27 | 676 |

with no tree growth control. Finally, we also considered a threshold ($\nu = 0.1$) for non-specificity measure. Higher values would allow for larger trees than those obtained with smaller thresholds.

Due to the disjunctive nature of the concepts represented by the inferred model [2], we have chosen to employ the *Dubois-Prade combination rule* [10] in order to pool BBAs. To compare inductive vs. deductive classification, the following metrics were computed:

- *match*: rate of the test cases (individuals) for which the inductive model and a reasoner predict the same membership (i.e. $+1 \mid +1$, $-1 \mid -1$, $0 \mid 0$);
- *commission*: rate of the cases for which predictions are opposite (i.e. $+1 \mid -1$, $-1 \mid +1$);
- *omission*: rate of test cases for which the inductive method cannot determine a definite membership ($-1, +1$) while the reasoner is able to do it;
- *induction*: rate of cases where the inductive method can predict a membership while it is not logically derivable.

### 4.2 Results

Tab. 2 shows the results of the experiments. A low commission rate is noticed for most of the ontologies, except BIOPAX. This rate is higher than the one observed with the standard algorithm. Besides, a low induction rate resulted but in the case of NTN.

In general, the proposed method returns more imprecise results than the results obtained with the TDTs [3], i.e. many times unknown-membership is assigned to test individuals. This is likely due to the combination rule employed. Indeed, the Dubois-Prade combination rule does not take into account the conflict [5]. The pooled BBA is obtained by combining other BBAs considering union of subset of the frame of discernment. Thus, the conflict does not exist and more imprecise results can be obtained. For example,this can occur, when we have two BBAs: the first one has $\{+1\}$ as the only focal element while the other one has $\{-1\}$ as the only focal element. The resulting BBA will have $\{-1, +1\}$ as focal element and an unknown case is returned.

With LUBM this phenomenon is very evident: there are not more induction cases, but the match and the omission rate are very high. In the case of NTN, when the DST-based method is employed, the match rate is lower than with the standard version and

[2] A concept can be obtained easily by visiting the tree and returning the conjunction of the concept descriptions encountered on a path.

**Table 2.** Results of the experiments using the original terminological trees (DLTree), the DST-TDTs induced with no growth control (DSTTree), and with a growth threshold (DSTG)

| *Ontology* | *Index* | *DLTree* | *DSTTree* | *DSTG* |
|---|---|---|---|---|
| FSM | *match* | 95.34±04.94 | 93.22±07.33 | 86.16±10.48 |
| | *commiss.* | 01.81±02.18 | 01.67±03.05 | 02.07±03.19 |
| | *omission* | 00.74±02.15 | 02.57±04.09 | 04.98±05.99 |
| | *induction* | 02.11±04.42 | 02.54±01.89 | 01.16±01.26 |
| LEO | *match* | 95.53±10.07 | 97.07±04.55 | 94.61±06.75 |
| | *commiss.* | 00.48±00.57 | 00.41±00.86 | 00.41±01.00 |
| | *omission* | 03.42±09.84 | 01.94±04.38 | 00.58±00.51 |
| | *induction* | 00.57±03.13 | 00.58±00.51 | 00.00±00.00 |
| LUBM | *match* | 20.78±00.11 | 79.23±00.11 | 79.22±00.12 |
| | *commiss.* | 00.00±00.00 | 00.00±00.00 | 00.00±00.00 |
| | *omission* | 00.00±00.00 | 20.77±00.11 | 20.78±00.12 |
| | *induction* | 79.22±00.11 | 00.00±00.00 | 00.00±00.00 |
| BioPax | *match* | 96.87±07.35 | 85.76±21.60 | 82.15±21.10 |
| | *commiss.* | 01.63±06.44 | 11.81±19.96 | 12.32±19.90 |
| | *omission* | 00.30±00.98 | 01.54±03.02 | 04.88±03.03 |
| | *induction* | 01.21±00.56 | 00.89±00.53 | 00.26±00.27 |
| NTN | *match* | 27.02±01.91 | 18.97±19.01 | 87.63±00.19 |
| | *commiss.* | 00.00±00.00 | 00.39±01.08 | 00.00±00.00 |
| | *omission* | 00.22±00.26 | 02.09±03.00 | 12.37±00.19 |
| | *induction* | 72.77±01.51 | 78.54±17.34 | 00.00±00.00 |

commission, omission and induction rate are higher. However, adding a tree-growth control threshold the method shows a more conservative behavior w.r.t. the first experimental condition. In the case of NTN, we observe a lower induction rate and higher match and omission rates. Instead, the commission error rate is lower. The increase of the induction rate and the decrease of the match rate for the DST-based method (without the tree-growth control threshold) are likely due to uncertain membership cases for which the algorithm can determine the class.

A final remark regards the stability for the proposed method: the outcomes show a higher standard deviation w.r.t. the original version, hence it seems less stable so far.

## 5 Conclusions and Extensions

In this work a novel type of terminological decision trees and the related learning algorithms have been proposed, in order to integrate forms of epistemic uncertainty in such an inductive classification model. We have shown how the DST can be employed together with machine learning methods for the Semantic Web representations as an alternative framework to cope with the inherent uncertainty and incompleteness. The proposed algorithm can discover potentially new (non logically derivable) assertions that can be used to complete the extensional part of a Web ontology (or a Linked Data dataset) whose expressiveness allows to represent concepts by means of disjunction and

complement operators. However, experimental results show that the current version of the method may have sometimes a worse performance, especially in terms of match rate and stability.

The proposed method can be extended along various directions. It is possible to use a total uncertainty measure that integrates conflicting evidence [9]. In the current version, we control the growth of the tree. A further extension may concern the definition of a pruning method based on the DST.

**Acknowledgments.** This work fulfils the research objectives of the PON 02_00563_34 70993 project "VINCENTE - A Virtual collective INtelligenCe ENvironment to develop sustainable Technology Entrepreneurship ecosystems" and the PON 02_00563 _3489339 project "PUGLIA@SERVICE - Internet-based Service Engineering enabling Smart Territory structural development" funded by the Italian Ministry of University and Research (MIUR).

## References

1. Rettinger, A., Lösch, U., Tresp, V., d'Amato, C., Fanizzi, N.: Mining the semantic web - statistical learning for next generation knowledge bases. Data Min. Knowl. Discov. 24, 613–662 (2012)
2. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.): The Description Logic Handbook, 2nd edn. Cambridge University Press (2007)
3. Fanizzi, N., d'Amato, C., Esposito, F.: Induction of concepts in web ontologies through terminological decision trees. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010, Part I. LNCS, vol. 6321, pp. 442–457. Springer, Heidelberg (2010)
4. Blockeel, H., De Raedt, L.: Top-down induction of first-order logical decision trees. Artif. Intell. 101, 285–297 (1998)
5. Klir, J.: Uncertainty and Information. Wiley (2006)
6. Rizzo, G., d'Amato, C., Fanizzi, N., Esposito, F.: Assertion prediction with ontologies through evidence combination. In: Bobillo, F., Costa, P.C.G., d'Amato, C., Fanizzi, N., Laskey, K.B., Laskey, K.J., Lukasiewicz, T., Nickles, M., Pool, M. (eds.) URSW 2008-2010/UniDL 2010. LNCS, vol. 7123, pp. 282–299. Springer, Heidelberg (2013)
7. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web. Morgan & Claypool Publishers (2011)
8. Denoeux, T.: A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Transactions on Systems, Man and Cybernetics 25, 804–813 (1995)
9. Sutton-Charani, N., Destercke, S., Denœux, T.: Classification trees based on belief functions. In: Denœux, T., Masson, M.-H. (eds.) Belief Functions: Theory & Appl. AISC, vol. 164, pp. 77–84. Springer, Heidelberg (2012)
10. Dubois, D., Prade, H.: On the combination of evidence in various mathematical frameworks. In: Flamm, J., Luisi, T. (eds.) Reliability Data Collection and Analysis. Eurocourses, vol. 3, pp. 213–241. Springer Netherlands (1992)

# Uncertainty in Ontology Matching: A Decision Rule-Based Approach

Amira Essaid[1,2], Arnaud Martin[2], Grégory Smits[2], and Boutheina Ben Yaghlane[3]

[1] LARODEC, University of Tunis, ISG Tunis, Tunisia
[2]IRISA, University of Rennes1, Lannion, France
[3] LARODEC, University of Carthage, IHEC Carthage, Tunisia

**Abstract.** Considering the high heterogeneity of the ontologies published on the web, ontology matching is a crucial issue whose aim is to establish links between an entity of a source ontology and one or several entities from a target ontology. Perfectible similarity measures, considered as sources of information, are combined to establish these links. The theory of belief functions is a powerful mathematical tool for combining such uncertain information. In this paper, we introduce a decision process based on a distance measure to identify the best possible matching entities for a given source entity.

**Keywords:** Theory of belief functions, decision rule, Jousselme distance, ontology matching.

## 1 Introduction

This paper proposes a decision rule based on a distance measure. This rule calculates the distance between a combined mass function and a categorical mass function and keep the hypotheses with the lowest distance. We propose this rule for its ability to give decision on composite hypotheses as well as its convenience to our domain of application, namely the semantic web and particularly the ontology matching where decision making is an important step.

Ontology matching is the process of finding for each entity of a source ontology $O_1$ its corresponding entity in a target ontology $O_2$. This process can focus on finding simple mappings (1:1) or complex mappings (1:n or n:1). The first consists in matching only one entity of $O_1$ with only one entity of $O_2$ whereas the second consists in finding either for one entity of $O_1$ its multiple correspondences of entities in $O_2$ or matching multiple entities of $O_1$ with only one entity of $O_2$. We are interested in this paper in finding simple mappings as well as the complex one of the form (1:n).

The matching process is performed through the application of matching techniques which are mainly based on the use of similarity measures. Since no similarity measure applied individually is able to give a perfect alignment, the exploitation of the complementarity of different similarity measures can yield

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 46–55, 2014.

to a better alignment. Combining these similarity measures may raise conflicts between the different results which should be modeled and resolved.

We suggest to use the theory of belief functions [4,12] as a tool for modeling the ontology matching and especially for combining the results of the different similarity measures. Due to the fact that we are working on an uncertain aspect and we are interested in finding complex matching which can be viewed as finding composite hypotheses formed from entities of two ontologies, we suggest to apply our proposed decision rule on the combined information and to choose for each entity of the source ontology, the entities of the target ontology with the lowest distance.

The remainder of this paper is organized as follows: we are interested in section 2 in defining the ontology matching process. In section 3, we recall the basic concepts underlying the theory of belief functions. In section 4, we present our decision rule based on a distance measure. Section 5 is devoted to the description of the credibilistic decision process for matching ontologies as well as the application of our proposed decision rule. Section 6 discusses an overview of some ontology matching approaches dealing with uncertainty. Finally, we conclude in section 7 and present future work.

## 2 Ontology Matching

The open nature of the semantic web [2] tends to encourage the development, for a domain of interest, of heterogeneous ontologies which differ from each other at the terminological level and/or the representational one. In order to mitigate the effect of semantic heterogeneity and to assure interoperability between applications that make use of these ontologies, a key challenge is to define an efficient and reliable *matching between ontologies* [7].

Formally, ontology matching is defined as a function A = f($O_1$, $O_2$, A', p, r). In fact, from a pair of ontologies to match $O_1$ and $O_2$, an input alignment $A'$, a set of parameters $p$, a set of oracles and resources $r$, the function $f$ returns an alignment $A$ between these ontologies. We note that parameters and resources refer to thresholds and external resources respectively.

With the new vision of the web that tends to make applications understandable by machines, an automatic and semi automatic discovery of correspondences between ontologies is required. The reader may refer to [7] for an exhaustive state of the art of ontology matching techniques.

## 3 The Theory of Belief Functions

### 3.1 Definitions

The frame of discernment $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ is a finite non empty set of $n$ elementary and mutually exclusive and exhaustive hypotheses related to a given problem. The power set of $\Theta$, denoted by $2^\Theta$ is defined as the set of singleton

hypotheses of $\Theta$, all possible disjunctions of these hypotheses as well as the empty set.

The basic belief assignment (*bba*) is the mapping from elements of the power set $2^{\Theta}$ onto $[0, 1]$ that satisfies:

$$m(\emptyset) = 0, \sum_{A \subseteq \Theta} m(A) = 1. \tag{1}$$

The value *m(A)* quantifies the part of belief exactly committed to the subset A of $\Theta$.

A focal element A is an element of $2^{\Theta}$ such that $m(A) \neq 0$.

From a given bba , the corresponding credibility and plausibility functions are respectively defined as:

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B). \tag{2}$$

and

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B). \tag{3}$$

The value *bel(A)* expresses the total belief that one allocates to A whereas the *pl(A)* quantifies the maximum amount of belief that might support a subset A of $\Theta$.

Some special *bbas* are defined in the theory of belief functions. Among them, the categorical bba which is a bba with a unique focal element different from the frame of discernment $\Theta$ and the empty set $\emptyset$, and which is defined as $m_X(X) = 1$.

### 3.2 Combination of Belief Functions

Let $S_1$ and $S_2$ be two distinct and independent sources providing two different bbas $m_1$ and $m_2$ defined on the same frame of discernment $\Theta$. These two bbas are combined by either the conjunctive rule of combination or the disjunctive rule.

– The conjunctive rule of combination is used when the two sources are fully reliable. This rule is defined in [14] as :

$$m_{1 \bigcirc\!\!\!\!\cap 2}(A) = \sum_{B \cap C = A} m_1(B) \times m_2(C). \tag{4}$$

The conjunctive rule can be seen as an unnormalized Dempster's rule of combination [4] which is defined by:

$$m_{1 \oplus 2}(A) = \begin{cases} \dfrac{\sum\limits_{B \cap C = A} m_1(B) \times m_2(C)}{1 - \sum\limits_{B \cap C = \emptyset} m_1(B) \times m_2(C)} & \forall A \subseteq \Theta,\ A \neq \emptyset \\ 0 & if\ A = \emptyset \end{cases} \tag{5}$$

The Dempster's rule of combination is normalized through $1 - \sum_{B \cap C = \emptyset} m_1(B) \times m_2(C)$ and it works under the closed world assumption where all the possible hypotheses of the studied problem are supposed to be enumerated on $\Theta$.

- The disjunctive rule is used when at least one of the sources is reliable without knowing which one of them. It is defined in [14] by:

$$m_{1 \textcircled{\cup} 2}(A) = \sum_{B \cup C = A} m_1(B) \times m_2(C). \tag{6}$$

### 3.3 Decision Making

Combining information provided by the different sources leads to a global one that has to be analyzed in order to choose the most likely hypothesis. Decision making can be held in two different ways.

- Decision on singletons: It means that the most likely solution to a given problem is one of the hypothesis of $\Theta$. To determine this most likely solution, one may:
  - Maximize the credibility: It consists on retaining the most credible hypothesis by giving the minimum of chances to each of the disjunctions.
  - Maximize the plausibility: It consists on retaining the most plausible hypothesis by giving the maximum of chances to each of the singletons.
  - Maximize the pignistic probability: It was introduced in [15] and it is the common used decision function because it represents a compromise between the maximum of credibility and the maximum of plausibility. The pignistic probability consists on choosing the most probable singleton hypothesis by dividing the mass attributed to each hypothesis, different from the singleton hypothesis, by the hypotheses composing it. It is given for each $A \in 2^{\Theta}$, $A \neq \emptyset$ by:

$$betP(X) = \sum_{A \in 2^{\Theta}, X \in A} \frac{m(A)}{|A|(1 - m(\emptyset))}. \tag{7}$$

    where $|A|$ represents the cardinality of A.
- Decision on unions of singletons: Few works were interested in making decision on composite hypotheses ([5], [1], [9]). The approach proposed in [1] helps to choose a solution of a given problem by considering all the elements contained in $2^{\Theta}$. This approach weights the decision functions listed previously by an utility function depending on the cardinality of the elements. For each A $\in 2^{\Theta}$ we have:

$$A = \underset{X \in 2^{\Theta}}{argmax}(m_d(X) pl(X)) \tag{8}$$

where $m_d$ is a mass defined by:

$$m_d(X) = K_d \lambda_X \left(\frac{1}{|X|^r}\right) \tag{9}$$

$r$ is a parameter in $[0, 1]$ for choosing a decision. When $r$ is equal to 0 it reflects a total indecision and when it is equal to 1 it means that we decide on a singleton. The value $\lambda_X$ is used to integrate the lack of knowledge about one of the elements X of $2^\Theta$. $K_d$ is a normalization factor.

## 4 Decision Rule Based on a Distance Measure

We aim in this paper to propose a decision rule helping us to choose the most likely hypothesis for a given problem after combining the information provided by different sources of information, i.e. bbas. This rule, based on a distance measure, is inspired from [13] and is defined as:

$$A = argmin(d(m, m_X)) \tag{10}$$

The proposed rule aims at calculating the distance between $m$ which is a combined bba (obtained after applying a combination rule) and $m_X$ is the categorical bba of X such that $X \in 2^\Theta$. The most likely hypothesis to choose is the hypothesis whose categorical bba is the nearest to the combined bba.

In order to make a decision:

- First, we have to identify the elements for which we have to construct the categorical bba. In fact, we choose to work on elements of $2^\Theta$ such that the cardinality of the element is less or equal to 2. This filtering is due to the fact that we want to limit the number of elements to be considered especially with a power set $2^\Theta$ of large cardinality.
- Then, we construct the categorical bba for each of the selected element.
- Finally, we calculate the distance between the combined bba and each of the categorical bbas. The minimum distance is kept and our decision corresponds to the categorical bba's element having the lowest distance with the combined bba.

For the calculation of the distance between the bbas, we use the Jousselme distance [8] which is specific to the theory of belief functions because of the matrix D defined on $2^\Theta$. This distance has the advantage to take into account the cardinality of the focal elements. This distance is defined for two bbas $m_1$ and $m_2$ as follows:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^t \underline{\underline{D}}(m_1 - m_2)} \tag{11}$$

where $\underline{\underline{\text{D}}}$ is a matrix based on Jaccard distance as a similarity measure between focal elements. This matrix is defined as:

$$D(A, B) = \begin{cases} 1 & \text{if A=B=}\emptyset \\ \frac{|A \cap B|}{|A \cup B|} & \forall A, B \in 2^\Theta \end{cases} \tag{12}$$

To illustrate the proposed decision rule, we take the following example. Let's consider the frame of discernment $\Theta = \{\theta_1, \theta_2, \theta_3\}$. The list of elements for which we have to construct their categorical bba are $\{\theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3\}$. Suppose that we have two sources $S_1$ and $S_2$ providing two different bbas $m_1$ and $m_2$ defined on the frame of discernment $\Theta$. The table 1 illustrates these two bbas as well as their combined bba obtained after applying the Dempster's rule of combination.

**Table 1.** bba1 and bba2 and their combined bba

| bba1 | bba2 | combined bba |
|---|---|---|
| $m_1(\theta_1) = 0.4$ | $m_2(\theta_2) = 0.2$ | $m_{comb}(\theta_1) = 0.3478$ |
| $m_1(\theta_2 \cup \theta_3) = 0.2$ | $m_2(\Theta) = 0.8$ | $m_{comb}(\theta_2) = 0.1304$ |
| $m_1(\Theta) = 0.4$ | | $m_{comb}(\Theta) = 0.3478$ |
| | | $m_{comb}(\theta_2 \cup \theta_3) = 0.1739$ |

The application of our proposed decision rule gives the results illustrated in table 2 where it shows for every element the distance obtained between the categorical bba of this element and the combined bba.

**Table 2.** Results of the proposed decision rule

| Element | Distance |
|---|---|
| $\theta_1$ | 0.537 |
| $\theta_2$ | 0.647 |
| $\theta_3$ | 0.741 |
| $\theta_1 \cup \theta_2$ | 0.472 |
| $\theta_1 \cup \theta_3$ | 0.536 |
| $\theta_2 \cup \theta_3$ | 0.529 |

Based on the results obtained in table 2, the most likely hypothesis to choose is the element $\theta_1 \cup \theta_2$.

## 5 Credibilistic Decision Process for Ontology Matching

In [6], we proposed a credibilistic decision process for ontology matching. In the following, we describe this process occurring mainly in three steps, then we will apply the proposed decision rule in order to find a correspondence for a given entity of the source ontology.

**1- Matching Ontologies:** We apply three name-based techniques (Levenshtein distance, Jaro distance and Hamming distance) for matching two ontologies $O_1$ and $O_2$ related to conference organization[1]. We have the following results:

[1] `http ://oaei.ontologymatching.org/2013/conference/index.html`

**Table 3.** Results of matching the entity ConferenceMember of $O_1$ with entities of $O_2$

| method | $e_2 \in O_2$ | $n$ |
|---|---|---|
| Levenshtein | $Conference\_fees$ | 0.687 |
| Jaro | $Conference$ | 0.516 |
| Hamming | $Conference$ | 0.625 |

This table shows that using the levenshtein distance, the entity *Conference-Member* matches to *Conference_fees* with a confidence value of 0.687.

**2- Modeling the Matching Under the Theory of Belief Functions:** We are interested here in modeling the matching results obtained in the previous step under the theory of belief functions.

- *Frame of discernment*: It contains all the entities of the target ontology $O_2$ for which a corresponding entity in the source ontology $O_1$ exists.
- *Source of information*: Every correspondence established by one of the matching techniques is considered as an information given by a source.
- *Basic Belief Assignments (bba)*: Once we get all the correspondences, we keep only those where an entity source $e_1 \in O_1$ has a correspondence when applying the three techniques. Then, we construct for each of the selected correspondence its mass function. The similarity measure obtained after applying a matching technique is interpreted as a mass. Due to the fact that for a source of information, the sum of mass functions has to be equal to 1, a mass will be affected to the total ignorance. Let's take the results illustrated in Table 3. In this table, we have information provided by three different sources respectively denoted by $S^{e_1}_{lev}$, $S^{e_1}_{jars}$ and $S^{e_1}_{hamming}$, where $e_1 = ConferenceMember$. The bba related to the source $S^{e_1}_{lev}$ is: $m_{S^{e_1}_{lev}}(Conference\_fees) = 0.687$ and $m_{S^{e_1}_{lev}}(\Theta) = 1 - 0.687 = 0.313$. The bbas for the other sources are constructed in the same manner.
- *Combination*: Let's resume the obtained bbas of the three sources. We have:
  - $m_{S^{e_1}_{lev}}(Conference\_fees) = 0.687$ and $m_{S^{e_1}_{lev}}(\Theta) = 0.313$
  - $m_{S^{e_1}_{jaro}}(Conference) = 0.516$ and $m_{S^{e_1}_{jaro}}(\Theta) = 0.484$
  - $m_{S^{e_1}_{hamming}}(Conference) = 0.625$ and $m_{S^{e_1}_{hamming}}(\Theta) = 0.375$

  Once we apply the Dempster's rule of combination, we obtain the following results:
  - $m^{e_1}_{comb}(Conference\_fees) = 0.2849$
  - $m^{e_1}_{comb}(Conference) = 0.5853$
  - $m^{e_1}_{comb}(\Theta) = 0.1298$

**3- Decision Making:** Based on the combined bba which takes into account all the information provided by the different sources, we will be able in this step to choose for each entity of the source ontology its corresponding in the

target ontology. For example, for the entity *ConferenceMember*, we will be able to decide if we have to match it with *Conference_fees* or *Conference* or simply we will not have a precise decision but rather an uncertain one where we can match *ConferenceMember* to $Conference\_fees \cup Conference$. We are interested in our credibilistic process to get an uncertain decision. For this purpose, we apply our proposed decision rule. First, we construct the categorical bba of elements having a cardinality equal to 2. For the example illustrated in figure 1 we have:

- $\mathrm{m}(conference\_document \cup conference) = 1$
- $\mathrm{m}(conference \cup conference\_fees) = 1$
- $\mathrm{m}(conference\_volume \cup committee) = 1$
- ...

Then we calculate the distance between the combined bba obtained previously and each of the categorical bba. Our best alignment corresponds to the nearest element to the combined bba in other words the element whose categorical bba has the minimum distance with the combined bba. For the entity *ConferenceMember* of the ontology $O_1$ we find $conference\_fees \cup conference$ with a distance equal to 0.52. This process is repeated for each entity of the source ontology in order to identify the most significant correspondences in the target ontology.

## 6 Related Works

Only few ontology matching methods have considered that dealing with uncertainty in a matching process is a crucial issue. We are interested in this section to present some of them where the probability theory [11] and the Dempster-Shafer theory ([3],[10],[16]) are the main mathematical models used. In [11], the authors proposed an approach for matching ontologies based on bayesian networks which is an extension of the BayesOWL. The BayesOWL consists in translating an OWL ontology into a bayesian network (BN) through the application of a set of rules and procedures. In order to match two ontologies, first the source and target ontologies are translated into $BN_1$ and $BN_2$ respectively. The mapping is processed between the two ontologies as an evidential reasoning between $BN_1$ and $BN_2$. The authors assume that the similarity information between a concept $C_1$ from a source ontology and a concept $C_2$ from a target ontology is measured by the joint probability distribution $\mathrm{P}(C_1, C_2)$.

In [3], the author viewed ontology matching as a decision making process that must be handled under uncertainty. He presented a generic framework that uses Dempster-Shafer theory as a mathematical model for representing uncertain mappings as well as combining the results of the different matchers. Given two ontologies $O_1$ and $O_2$, the frame of discernment represents the Cartesian product e x $O_2$ where each hypothesis is the couple $< e, e_i >$ such as e $\in O_1$ and $e_i \in O_2$. Each matcher is considered as an expert that returns a similarity measure converted into a basic belief mass. The Dempster rule of combination is used to combine the results provided by a matcher. The pairs with plausibility

and belief below a given threshold are discarded. The remaining pairs represent the best mapping for a given entity.

Although, the authors in [10] handle uncertainty in the matching process, their proposal differ from that proposed in [3]. In fact, they use the Dempster-Shafer theory in a specific context of question answering where including uncertainty may yield to better results. Not like in [3], they did not give in depth how the frame of discernment is constructed. In addition to that, uncertainty is handled only once the matching is processed. In fact, the similarity matrix is constructed for each matcher. Based on this matrix, the results are modeled using the theory of belief functions and then they are combined.

In [16], the authors focused on integrating uncertainty when matching ontologies. The proposed method modeled and combined the outputs of three ontology matchers. For an entity e $\in O_1$, the frame of discernment $\Theta$ is composed of mappings between e and all the concepts in an ontology $O_2$. The different similarity values obtained through the application of the three matchers are interpreted as mass values. Then, a combination of the results of the three matchers is performed.

## 7 Conclusion and Perspectives

In this paper, we proposed a decision rule based on a distance measure. This decision rule helps to choose the most likely hypothesis for a given problem. It is based on the calculation of the distance between a combined bba and a categorical bba. We apply this rule in our proposed credibilistic decision process for the ontology matching. First, we match two ontologies. Then, the obtained correspondences are modeled under the theory of belief functions. Based on the obtained results, a decision making is performed by applying our proposed decision rule.

In the future, we aim at applying other matching techniques. We are interested also in constructing an uncertain ontology based on the obtained results after a decision making and handling experimentations to qualitatively assess the relevance of our approach.

## References

1. Appriou, A.: Approche Générique de la Gestion de l'Incertain dans les Processus de Fusion Multisenseur. Traitement du Signal 22, 307–319 (2005)
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American 284(5), 34–43 (2001)
3. Besana, P.: A Framework for Combining Ontology and Schema Matchers with Dempster Shafer. In: International Semantic Web Conference. LNCS, vol. 4273, pp. 196–200. Springer, HeidelBerg (2006)
4. Dempster, A.P.: Upper and Lower Probabilities Induced by a Multivalued Mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
5. Denoeux, T.: Analysis of Evidence-theoretic Decision Rules for Pattern Classification. Pattern Recognition 30(7), 1095–1107 (1997)

6. Essaid, A., Martin, A., Smits, G., Ben Yaghlane, B.: Processus de Décision Crédibiliste pour l'Alignement des Ontologies. LFA, Reims, France (2013)
7. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer (2007)
8. Jousselme, A.L., Grenier, D., Bossé, E.: A New Distance Between Two Bodies of Evidence. Information Fusion 2, 91–101 (2001)
9. Martin, A., Quidu, I.: Decision Support with Belief Functions Theory for Seabed Characterization. In: International Conference on Information Fusion, Cologne, Germany (2008)
10. Nagy, M., Vargas-Vera, M., Motta, E.: DSSIM-Ontology mapping with uncertainty. In: International Semantic Web Conference. LNCS, vol. 4273, Springer, HeidelBerg (2006)
11. Pan, R., Ding, Z., Yu, Y., Peng, Y.: A bayesian network approach to ontology mapping. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 563–577. Springer, Heidelberg (2005)
12. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
13. Smarandache, F., Martin, A., Osswald, C.: Contradiction Measures and Specificity Degrees of Basic Belief Assignments. In: International Conference on Information Fusion, Chicago, USA (2011)
14. Smets, P.: The Combination of Evidence in the Transferable Belief Model. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(5), 447–458 (1990)
15. Smets, P.: Decision Making in the TBM: the Necessity of the Pignistic Transformation. International Journal of Approximate Reasonning 38, 133–147 (2005)
16. Wang, Y., Liu, W., Bell, D.: Combining Uncertain Outputs from Multiple Ontology Matchers. In: Prade, H., Subrahmanian, V.S. (eds.) SUM 2007. LNCS (LNAI), vol. 4772, pp. 201–214. Springer, Heidelberg (2007)

# A Practical Application of Argumentation in French Agrifood Chains

Madalina Croitoru, Rallou Thomopoulos, and Nouredine Tamani

LIRMM (University of Montpellier II & CNRS), INRIA Sophia-Antipolis, France

**Abstract.** Evaluating food quality is a complex process since it relies on numerous criteria historically grouped into four main types: nutritional, sensorial, practical and hygienic qualities. They may be completed by other emerging preoccupations such as the environmental impact, economic phenomena, etc. However, all these aspects of quality and their various components are not always compatible and their simultaneous improvement is a problem that sometimes has no obvious solution, which corresponds to a real issue for decision making. This paper proposes a decision support method guided by the objectives defined for the end products of an agrifood chain. It is materialized by a backward chaining approach based on argumentation.

## 1 Introduction

In agrifood chains, the products traditionally go through the intermediate stages of processing, storage, transport, packaging and reach the consumer (the demand) from the producer (the supply). More recently, due to an increase in quality constraints, several parties are involved in the production process, such as consumers, industrials, health and sanitary authorities, expressing their requirements on the final product as different points of view which could be conflicting. The notion of reverse engineering control, in which the demand sets the specifications of desired products and it is up to the supply to adapt and find its production requirements to respond, can be considered in this case.

In this article, we discuss two aspects of this problem. First, we accept the idea that specifications cannot be established and several complementary points of view - possibly contradictory - can be expressed (nutritional, environmental, taste, etc.). We then need to assess their compatibility (or incompatibility) and identify solutions satisfying a maximum set of viewpoints. To this end we proposed a logical framework based on argumentation and introduced a method of decision making based on backward chaining for the bread industry. This method detects inconsistencies and proposes several options to solve the problem.

Since a joint argumentation - decision support approach is highly relevant to the food sector [24], the **contribution** of the paper is **to present a real use case of an argumentation process in the agrifood domain**. For technical details (such as the way we introduce the notion of viewpoint / goal in this setting based on the notion of backwards chaining reasoning and show how to use those

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 56–66, 2014.

techniques in a concrete application) please check the technical report [11] which complements this paper in terms of technical results. **In this paper we aim at solely presenting the practical application of argumentation and how it can benefit the agri-food domain.** In Section 2, we introduce the real scenario considered in the application. In Section 3, we motivate our technical and modeling choices. In Section 4, the developed approach is introduced. It relies on an instantiation of a logic based argumentation framework based on a specific fragment of first order logic. In Section 5, we explain how to ensure the soundness and completeness of our agronomy application method. In Section 6, some evaluation results are presented. Finally, Section 7 concludes the paper.

## 2 Scenario

The case of study considered in this paper relates to the debate around the change of ash content in flour used for common French bread. Various actors of the agronomy sector are concerned, in particular the Ministry for Health through its recommendations within the framework of the PNNS ("National Program for Nutrition and Health"), the millers, the bakers, the nutritionists and the consumers. The PNNS recommends to privilege the whole-grain cereal products and in particular to pass to a common bread of T80 type, i.e made with flour containing an ash content (mineral matter rate) of 0.8%, instead of the type T65 (0.65% of mineral matter) currently used. Increasing the ash content comes down to using a more complete flour, since mineral matter is concentrated in the peripheral layers of the wheat grain, as well as a good amount of components of nutritional interest (vitamins, fibres). However, the peripheral layers of the grain are also exposed to the phytosanitary products, which does not make them advisable from a health point of view, unless one uses organic flour. Other arguments (and of various nature) are in favour or discredit whole-grain bread. From an organoleptic point of view for example, the bread loses out in its "being crusty". From a nutritional point of view, the argument according to which the fibres are beneficial for health is discussed, some fibres could irritate the digestive system. From an economic point of view, the bakers fear selling less bread, because whole-grain bread increases satiety – which is beneficial from a nutritional point of view, for the regulation of the appetite and the fight against food imbalances and pathologies. However whole-grain bread requires also less flour and more water for its production, thus reducing the cost. The millers also fear a decrease in the quality of the technical methods used in the flour production.

*Beyond the polemic on the choice between two alternatives (T65 or T80), one can take the debate further by distinguishing the various points of view concerned, identifying the desirable target characteristics, estimating the means of reaching that point.* **The contribution of this paper is showing how using argumentation can help towards such practical goals.**

## 3 Motivation

In this paper we will elicit the points of view and the desirable target characteristics by the means of interviews with agronomy experts. Once the target characteristics identified, finding the means of reaching them will be done automatically by a combination of reverse engineering and argumentation. *The reverse engineering will be used in order to find the complete set of actions to take towards a given characteristic, for all characteristics. In certain cases the actions to take will be inconsistent. Argumentation will then be employed in order to identify actions that can be accepted together.*

### 3.1 Reverse Engineering

While reverse engineering has been widely employed in other Computer Science domains such as multi-agent systems or requirements engineering, it is quite a novel methodology when applied in agronomy. In agrifood chains, the products traditionally go through the intermediate stages of processing, storage, transport, packaging and reach the consumer (the demand) from the producer (the supply). It is only recently, due to an increase in quality constraints, that the notion of reverse engineering control has emerged. In this case the demand (and not the supply) sets the specifications of desired products and it is up to the supply to adapt and find its ways to respond. In what follows, starting from the desired target criteria for the final product, the methods allowing one to identify ways to achieve these criteria (by intervention on the various stages of the supply chain) are named "reverse engineering".

Reverse engineering is known to be challenging from a methodological viewpoint. This is due to two main aspects. First, the difficulty of defining the specifications for the expected finished product. The desired quality criteria are multiple, questionable, and not necessarily compatible. The second difficulty lies in the fact that the impact of different steps of food processing and their order is not completely known. Some steps are more studied than others, several successive steps can have opposite effects (or unknown effects), the target criteria may be outside of the characteristics of products. Moreover, reconciling different viewpoints involved in the food sector still raises unaddressed issues. The problem does not simply consist in addressing a multi-criteria optimisation problem [7]: the domain experts would need to be able to justify why a certain decision (or set of possible decisions) is taken.

### 3.2 Argumentation

Argumentation theory in general [16] is actively studied in the literature, some approaches combining argumentation and multi criteria decision making [1].

**Logic-Based Argumentation.** In this paper we present a methodology combining reverse engineering and logical based argumentation for selecting the actions to take towards the agronomy application at hand. The logical instantiation

language is a subset of first order logic denoted in this paper $\mathcal{SRC}$ equivalent to Datalog+- [9], Conceptual Graphs or Description Logics (more precisely the $\mathcal{EL}$ fragment [3] and DL-Lite families [10]). All above mentioned languages are logically equivalent in terms of representation or reasoning power. The reason why this application is using $\mathcal{SRC}$ is the graph based representation proper to $\mathcal{SRC}$ (and not to the other languages). This graph based representation (implemented in the Cogui tool) makes the language suitable for interacting with non computing experts [20]. More on the visual appeal of Cogui for knowledge representation and reasoning can be found in [20]. In the following we use the instantiation of [12] for defining what an argument and an attack is.

## 4 Approach

In a nutshell our methodology is as follows. The set of goals, viewpoints as well as the knowledge associated with the goals / viewpoints is elicited either by the means of interviews with the domain experts or manually from different scientific papers. This step of the application is the most time consuming but the most important. If the elicited knowledge is not complete, sound or precise the outcome of the system is compromised. Then, based on the knowledge elicited from the knowledge experts and the goals of the experts, we enrich the knowledge bases using reverse engineering (implemented using backwards chaining algorithms). Putting together the enriched knowledge bases obtained by backwards chaining from the different goals will lead to inconsistencies. The argumentation process is used at this step and the extensions yield by the applications are computed. Based on the extensions and the associated viewpoints we can use voting functions to determine the application choice of viewpoints.

### 4.1 Use Case Real Data

Expressing the target characteristics – or goals – according to various points of view consists of identifying the facets involved in the construction of product quality: points of view, topics of concern such as nutrition, environment, technology, etc. In addition, such viewpoints have to be addressed according to their various components (fibres, minerals, vitamins, etc). Desirable directions need to be laid down, and we first consider them independent.

The considered sources of information include, from most formal to less formal: (1) peer reviewed scientific papers; (2) technical reports or information posted on websites; (3) conferences and scientific meetings around research projects; (4) expert knowledge obtained through interviews. The scientific articles we have analysed include: [6,23,15,19]. [6] compares the different types of flour from a nutritional point of view. [23] explores the link between fibre and satiety. [15] deals with consumer behaviour and willingness to pay. They focus on French baguette when information concerning the level of fibres is provided, and they base their results on statistical studies of consumer panels. [19] provides a summary of the nutritional aspects of consumption of bread and the link with technological aspects.

We also reviewed technical reports available on official websites on health policy: the public PNNS (National Program for Nutrition and Health) [21,22], the European project Healthgrain (looking at improving nutrition and health through grains) [13,18], as well as projects and symposia on sanitary measures regarding the nutritional, technological and organoleptic properties of breads [14,8,2,17]. Finally, several interviews were conducted to collect domain expert knowledge, in particular technology specialists in our laboratory.

A summary of the results obtained in the baking industry is synthesised in Figure 1 regarding the nutritional and organoleptic aspects.

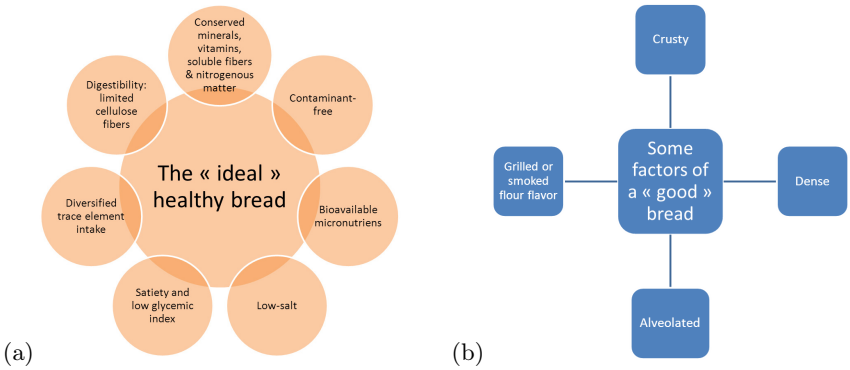



**Fig. 1.** Nutritional (a) and organoleptic (b) goals

## 5 Technical Soundness

In this section we explain the technical results that ensure the soundness and completeness of our agronomy application method. The section is composed of three parts. A first subsection explains the logical subset of first order logic language employed in the paper. The second subsection deals with arguments and attacks and how to obtain extensions when a knowledge base expressed under this language is inconsistent. Last, the third section shows how we used reverse engineering to complete the knowledge base with all possible actions and how argumentation can be used in order to select consistent subsets of knowledge which support given actions.

### 5.1 The Logical Language

In the following, we give the general setting knowledge representation language used throughout the paper.

A knowledge base is a 3-tuple $\mathcal{K} = (\mathcal{F}, \mathcal{R}, \mathcal{N})$ composed of three finite sets of formulae: a set $\mathcal{F}$ of facts, a set $\mathcal{R}$ of rules and a set $\mathcal{N}$ of constraints. Please check the technical report [11] for more formal details on the elements of the

language. For space reasons, in the following we will simply give an intuition of the expressivity of this language by the means an example. We thus prefer to explain in detail the application contribution of the paper and keep the technical details fully accessible in the technical report [11].

Let $\mathcal{K} = (\mathcal{F}, \mathcal{R}, \mathcal{N})$ where :

- $\mathcal{F} = \{F_1\} = \{CurrentExtractionRate(T65)\}$
- $\mathcal{R}$ contains the following rules:
  - $R_1 = \forall\ x,y\ (Bread(x) \wedge ExtractionRate(y,x) \wedge Decrease(y) \rightarrow Digestible(x))$
  - $R_2 = \forall\ x,z\ (Bread(x) \wedge SaltAdjunction(z,x) \wedge Decrease(z) \rightarrow LowSalt(x))$
  - $R_3 = \forall\ x,y\ (Bread(x) \wedge ExtractionRate(y,x) \wedge Growth(y) \rightarrow TraceElementRich(x))$
  - $R_4 = \forall\ x,y\ (Bread(x) \wedge ExtractionRate(y,x) \wedge Decrease(y) \rightarrow PesticideFree(x))$
- $\mathcal{N}$ contains the following negative constraint:
  - $N = \neg(\exists\ x\ (Growth(x) \wedge Decrease(x)))$

Reasoning consists of applying rules on the set $\mathcal{F}$ and thus inferring new knowledge. A rule $R = (H, C)$ is *applicable* to set $\mathcal{F}$ if and only if there exists $\mathcal{F}' \subseteq \mathcal{F}$ such that there is a homomorphism $\sigma$ from the hypothesis of $\mathcal{R}$ to the conjunction of elements of $\mathcal{F}'$. The rule application can be done using forward chaining (denoted $R(F)$) or backwards chaining (denoted $R^{-1}(F)$). A set $\{F_1, \ldots, F_k\} \subseteq \mathcal{F}$ is $\mathcal{R}$-*inconsistent* if and only if there exists a constraint $N \in \mathcal{N}$ such that the closure of facts with the rules entails the negative constraint.

### 5.2 Arguments and Attacks

This section shows that it is possible to define an instantiation of Dung's abstract argumentation theory [16] that can be used to reason with an inconsistent ontological KB. The notion of argument and attack used here follow the one introduced by [12]. As previously explained we simply show an example of argument and (asymmetric) attack. Let us introduce some argumentation notions.

Given a knowledge base $\mathcal{K} = (\mathcal{F}, \mathcal{R}, \mathcal{N})$, the *corresponding argumentation framework* $\mathcal{AF}_K$ is a pair $(\mathcal{A} = \texttt{Arg}(\mathcal{F}), \texttt{Att})$ where $\texttt{Arg}(\mathcal{F})$ is the set of all arguments that can be constructed from $\mathcal{F}$ and $\texttt{Att}$ is the corresponding attack relation as specified in [12]. Let $\mathcal{E} \subseteq \mathcal{A}$ and $a \in \mathcal{A}$. We say that $\mathcal{E}$ is *conflict free* iff there exists no arguments $a, b \in \mathcal{E}$ such that $(a, b) \in Att$. $\mathcal{E}$ *defends* $a$ iff for every argument $b \in \mathcal{A}$, if we have $(b, a) \in Att$ then there exists $c \in \mathcal{E}$ such that $(c, b) \in Att$. $\mathcal{E}$ is *admissible* iff it is conflict free and defends all its arguments. $\mathcal{E}$ is a *complete extension* iff $\mathcal{E}$ is an admissible set which contains all the arguments it defends. $\mathcal{E}$ is a *preferred extension* iff it is a maximal admissible set (with respect to set inclusion). $\mathcal{E}$ is a *stable extension* iff it is conflict-free and for all $a \in \mathcal{A} \setminus \mathcal{E}$, there exists an argument $b \in \mathcal{E}$ such that $(b, a) \in Att$. $\mathcal{E}$ is a *grounded extension* iff $\mathcal{E}$ is a minimal (for set inclusion) complete extension. An argument is sceptically accepted if it is in all extensions, credulously accepted if it is in at least one extension and rejected if it is not in any extension.

### 5.3 Formalising the Use Case

In this subsection we illustrate the applicative use of the argumentation notions in a goal-based inconsistent knowledge base obtained from an agri-food scenario.

Let the knowledge base defined in section 5.1 and the goal set $\mathcal{G}$ as:

- $G_1 = \exists\ p\ (Bread(p) \wedge Digestible(p))$, where $\kappa(G_1) = nutrition$
- $G_2 = \exists\ p\ (Bread(p) \wedge LowSalt(p))$, where $\kappa(G_2) = nutrition$
- $G_3 = \exists\ p\ (Bread(p) \wedge TraceElementRich(p))$, where $\kappa(G_3) = nutrition$
- $G_4 = \exists\ p\ (Bread(p) \wedge PesticideFree(p))$, where $\kappa(G_4) = sanitary$.

Then:

- $\mathcal{K}_1 = (\mathcal{F}_1, \mathcal{R}, \mathcal{N})$ where $\mathcal{F}_1 = \mathcal{F} \cup \mathcal{R}^{-1}(G_1)$ contains the following facts:
  - $F_1 = CurrentExtractionRate(T65)$
  - $F_2 = Bread(p) \wedge ExtractionRate(\tau,p) \wedge Decrease(\tau)$
- $\mathcal{K}_2 = (\mathcal{F}_2, \mathcal{R}, \mathcal{N})$ where $\mathcal{F}_2 = \mathcal{F} \cup \mathcal{R}^{-1}(G_2)$ contains the following facts:
  - $F_1 = CurrentExtractionRate(T65)$
  - $F_3 = Bread(p) \wedge SaltAdjunction(s,p) \wedge Decrease(s)$
- $\mathcal{K}_3 = (\mathcal{F}_3, \mathcal{R}, \mathcal{N})$ where $\mathcal{F}_3 = \mathcal{F} \cup \mathcal{R}^{-1}(G_3)$ contains the following facts:
  - $F_1 = CurrentExtractionRate(T65)$
  - $F_4 = Bread(p) \wedge ExtractionRate(\tau,p) \wedge Growth(\tau)$
- $\mathcal{K}_4 = (\mathcal{F}_4, \mathcal{R}, \mathcal{N})$ where $\mathcal{F}_4 = \mathcal{F} \cup \mathcal{R}^{-1}(G_4)$ contains the following facts:
  - $F_1 = CurrentExtractionRate(T65)$
  - $F_2 = Bread(p) \wedge ExtractionRate(\tau,p) \wedge Decrease(\tau)$

  Finally $\mathcal{K}_{agg} = (\mathcal{F} \bigcup_{i=1,\ldots,n} \mathcal{R}^{-1}(G_i), \mathcal{R}, \mathcal{N})$ where $\mathcal{F} \bigcup_{i=1,\ldots,n} \mathcal{R}^{-1}(G_i) = \{F_1, F_2, F_3, F_4\}$.

As observed in the previous example, it may happen that $\mathcal{K}_{agg}$ is inconsistent (and it does so even for goals belonging to the same viewpoint). We then use argumentation, which, by the means of extensions will isolate subsets of facts we can accept together (called extensions). Furthermore, the extensions will allow us to see which are the viewpoints associated to each maximal consistent subset of knowledge (by the means of the function $\kappa$). Once we obtain this we can either use simple voting procedures to find out which viewpoint to follow or other preference based selection.

The argument framework we can construct from the above knowledge base is $(\mathcal{A}, Att)$ where $\mathcal{A}$ contains the following:

- $a = (\{F_2\}, F_2, R_1(F_2))$ where $R_1(F_2) = Bread(p) \wedge ExtractionRate(\tau,p) \wedge Decrease(\tau) \wedge Digestible(p)$.
- $b = (\{F_4\}, F_4, R_3(F_4))$ where $R_3(F_4) = Bread(p) \wedge ExtractionRate(\tau,p) \wedge Growth(\tau) \wedge TraceElementRich(p)$.
- $c = (\{F_2\}, F_2, R_4(F_2))$ where $R_4(F_2) = Bread(p) \wedge ExtractionRate(\tau,p) \wedge Decrease(\tau) \wedge PesticideFree(p)$.

• $d = (\{F_3\}, F_3, R_2(F_3))$ where $R_2(F_3) = \textit{Bread(p)} \wedge \textit{SaltAdjunction(s,p)} \wedge \textit{Decrease(s)} \wedge \textit{LowSalt(p)}$ and
$Att = \{(a, b), (b, a), (b, c), (c, b)\}$.

In the defined argumentation system, we can now derive:

• $Ext_{stable}(\mathcal{A}, Att) = Ext_{semi-stable}(\mathcal{A}, Att) = Ext_{preferred}(\mathcal{A}, Att) = \{\{a, c, d\}, \{b, d\}\}$.

Starting from the extensions $Ext_x(\mathcal{A}, Att)$, the proposed decision support system functions as follows: for every extension $\varepsilon \in Ext_x(\mathcal{A}, Att)$ :

- Consider the facts occurring in the arguments of $\varepsilon$ ;
- Identify the knowledge bases $\mathcal{K}_i$ where these facts occur;
- Obtain the goals $G_i$ which are satisfied by the extension;
- Using the $\kappa$ function to obtain the viewpoints corresponding to these goals;
- Show domain experts the set of goals, and compatible viewpoints corresponding to the given extension.

This method allows us to obtain a set of options equal to the cardinality of $Ext_x(\mathcal{A}, Att)$. To make a final decision several possibilities can be considered and presented to the experts:

- Maximise the number of goals satisfied;
- Maximise the number of viewpoints satisfied;
- Use preference relations of experts on goals and / or viewpoints.

In the previous example (please recall that the goals $G_1$ and $G_2$ are associated with the nutritional viewpoint while $G_4$ is associated with the sanitary viewpoint) we have:

- The first extension $\{a, c, d\}$ is based on the facts $F_2$ and $F_3$ obtained from $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_4$ that satisfy the goals $G_1$, $G_2$ and $G_4$.
- The second extension $\{b, d\}$ is based on $F_3$ and $F_4$ obtained from $\mathcal{K}_2$ and $\mathcal{K}_3$ satisfying $G_2$ and $G_3$ both associated with the nutritional viewpoint.

One first possibility (corresponding to the extension $\{a, c, d\}$) consists of accomplishing $F_2$ and $F_3$ and allows to satisfy the biggest number of goals and viewpoints.

The second possibility (corresponding to the extension $\{b, d\}$) consists of accomplishing $F_3$ and $F_4$. It would satisfy two goals and one viewpoint. It could be considered though if the goal $G_3$ (not satisfied by the first option) is preferred to the others.

## 6 Evaluation

The evaluation of the implemented system was done via a series of interviews with domain experts. The above knowledge and reasoning procedures were implemented using the Cogui knowledge representation tool [20], with an extension of 2000 lines of supplemental code. Three experts have validated our approach:

two researchers in food science and cereal technologies of the French national institute of agronomic research, specialists respectively of the grain-to-flour transformation process and of the breadmaking process, and one industrial expert - the president of the French National Institute of Bread and Pastry.

The first meeting dealt with the delimitation of the project objectives and addressed fundamental questions such as: Is it possible to uniquely define a "good" bread? Which scenarii of "good bread" should be considered? How could they be defined from a nutritional, sanitary, sensorial and economic point of view? Which are the main known ways to achieve them? Then a series of individual interviews constituted the elicitation phase. Each expert gave more arguments which were complementing one another.

In the following plenary meeting the real potential of the approach was shown. The experts were formulating goals and viewpoints they were interested in and the Cogui system together with the argumentation extension was yielding the associated possible propositions.

Two interests of the approach were more particularly highlighted. They concern cognitive considerations. Firstly, experts were conscious that the elicitation procedure was done according to their thought processes, that is, in a forward way which is more natural and intuitive. The system was thus able to restitute the knowledge in a different manner than the experts usually do. Secondly, from a problem that could initially seem simple, the experts realized that it covered a huge complexity that a human mind could hardly address on its own. The tool is currently available to them under restricted access.

## 7 Conclusion

Even if argumentation based decision making methods applied to the food industry were also proposed by [4,5], this paper addresses a key issue in the context of current techniques used by the food sector and namely addressing reverse engineering. Moreover, in this approach, an argument is used as a method computing compatible objectives in the sector. This case study represents an original application and an introspective approach in the agronomy field by providing an argumentation based decision-support system for the various food sectors. It requires nevertheless the very expensive task of knowledge modeling. Such task, in its current state cannot be automated. It strongly depends on the quality of expert opinion and elicitation (exhaustiveness, certainty, etc). The current trend for decision-making tools includes more and more methods of argumentation as means of including experts in the task of modeling and the decision-making processes. Another element to take into account, not discussed in this paper, is the difficulty of technologically (from an agronomy viewpoint) putting in place the facts of each option. Modeling this aspect in the formalism has still to be studied.

## References


1. Amgoud, L., Prade, H.: Using arguments for making and explaining decisions. Artificial Intelligence 173(3-4), 413–436 (2009)
2. AQUANUP (2009), http://www.inra.fr/inra_cepia/vous_recherchez/des_projets/france/aquanup
3. Baader, F., Brandt, S., Lutz, C.: Pushing the el envelope. In: Proc. of IJCAI 2005 (2005)
4. Bourguet, J.-R.: Contribution aux methodes d'argumentation pour la prise de decision. Application a l'arbitrage au sein de la filiere cerealiere. Thèse de doctorat, Université Montpellier II, Montpellier, France (2010)
5. Bourguet, J.-R., Thomopoulos, R., Mugnier, M.-L., Abécassis, J.: An artificial intelligence-based approach to deal with argumentation applied to food quality in a public health policy. Accepted for Publication in: Expert Systems With Applications (2013)
6. Bourre, J.-M., Bégat, A., Leroux, M.-C., Mousques-Cami, V., Pérandel, N., Souply, F.: Valeur nutritionnelle (macro et micro-nutriments) de farines et pains français. Médecine et Nutrition 44(2), 49–76 (2008)
7. Bouyssou, D., Dubois, D., Pirlot, M., Prade, H.: Decision-making process – Concepts and Methods. Wiley (2009)
8. CADINNO. Information, choix, consommateurs responsables: des leviers pour un développement durable? (2008), http://www.melissa.ens-cachan.fr/IMG/pdf/Colloque_CadInno_FR.pdf
9. Calì, A., Gottlob, G., Lukasiewicz, T., Marnette, B., Pieris, A.: Datalog+/-: A family of logical knowledge representation and query languages for new applications. In: LICS, pp. 228–242 (2010)
10. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The dl-lite family. J. Autom. Reasoning 39(3), 385–429 (2007)
11. Croitoru, M., Thomopoulos, R., Tamani, N.: Using argumentation in a french agrifood chain application: Technical report. Technical report, GrapIK, INRIA, University of Montpellier 2 (December 2013)
12. Croitoru, M., Vesic, S.: What can argumentation do for inconsistent ontology query answering? In: Liu, W., Subrahmanian, V.S., Wijsen, J. (eds.) SUM 2013. LNCS, vol. 8078, pp. 15–29. Springer, Heidelberg (2013)
13. Dean, M., Sheperd, R., Arvola, A., Lampila, P., Lahteenmaki, L., Vassalo, M., Saba, A., Claupein, E., Winkelmann, M.: Report on consumer expectations of health benefits of modified cereal products. Technical report, University of Surrey, UK (2007)
14. DINABIO. Proceedings of dinabio développement et innovation en agriculture biologique (2008), http://www.inra.fr/ciag/revue_innovations_agronomiques/volume_4_janvier_2009
15. Dubuisson-Quellier, S.: De la routine à la délibération. les arbitrages des consommateurs en situation d'achat. Réseaux 135/136, 253–284 (2006)
16. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and $n$-person games. Artificial Intelligence Journal 77, 321–357 (1995)
17. FCN. Fibres, céréales et nutrition (2009), http://www.inra.fr/content/view/full/24670029,
18. HEALTHGRAIN (2009), http://www.healthgrain.org

19. Layat, T.: Place du pain dans l'équilibre alimentaire. Pratiques en Nutrition 7(26), 45–50 (2011)
20. Michel, C., Mugnier, M.-L., Croitoru, M.: Visual reasoning with graph-based mechanisms: the good, the better and the best. The Knowledge Engineering Review 28, 249–271 (2013)
21. PNNS (documents statutaires) (2010), http://www.sante.gouv.fr/htm/pointsur/nutrition/pol_nutri4.htm
22. PNNS (site web) (2010), http://www.mangerbouger.fr/menu-secondaire/pnns/le-pnns
23. Slavin, J., Green, H.: Diatery fibre and satiety. British Nutrition Foundation 32(1), 32–42 (2007)
24. Thomopoulos, R., Charnomordic, B., Cuq, B., Abécassis, J.: Artificial intelligence-based decision support system to manage quality of durum wheat products. Quality Assurance and Safety of Crops & Foods 1(3), 179–190 (2009)

# Soft Fusion of Heterogeneous Image Time Series

Mar Bisquert[1,3,*], Gloria Bordogna[2], Mirco Boschetti[2],
Pascal Poncelet[1], and Maguelonne Teisseire[3]

[1] LIRMM, UM2, Montpellier, France
pascal.poncelet@lirmm.fr
[2] CNR IREA, Milano, Italy
{bordogna.g,boschetti.m}@irea.cnr.it
[3] IRSTEA, UMR TETIS. Montpellier, France
{mar.bisquert-perles,maguelonne.teisseire}@teledetection.fr

**Abstract.** In this contribution we analyze the problem of the fusion of time series of heterogeneous remote sensing images to serve classification and monitoring activities which can aid farming applications such as crop classification, change detection and monitoring. We propose several soft fusion operators that are based on different assumptions and model distinct desired properties. Conducted experiments on various geographic regions have been carried out and illustrate the effectiveness of our proposal.

## 1 Introduction

Fusion techniques in remote sensing may be useful for obtaining dense time series of high resolution images. Low resolution images use to have high temporal frequency while they have limited spatial information. Conversely, even if they have higher economical costs, high resolution images may have lower temporal frequency but obviously they provide higher spatial resolution. Fusion methods between high and low resolution images can be applied for simulating detailed images in dates where they are not available. Having a dense temporal series of high resolution images is important in numerous studies including classification, monitoring, change detection, etc. In this sense, image fusion is the combination of two or more images of the same scene, taken by different sensors at either the same or subsequent time instants, into a synthesized image that is more informative and more suitable for a given task, such as for visual perception, or computer processing [1], i.e., conveying information not previously available [2].

Image fusion can be performed at distinct representation levels of the information in the input images. When performed at pixel level, i.e.,on a pixel-by-pixel basis, as in our case, it serves the purpose to generate a fused image in which the information associated with a pixel, is determined from the coreferred input pixels in the source images to improve the performance of image processing tasks, such as segmentation, classification or change detection.

Fuzzy set theory has been indicated as a suitable framework for modeling image soft fusion since it allows representing the vague and often heuristic fusion criteria.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 67–76, 2014.

For instance, in [3,4] image fusion has been modeled by a fuzzy rule-based approach, in which the expert's vague criteria are represented by fuzzy rules.

Another fuzzy approach is the one proposed in [5] where soft data fusion is regarded as a process in which mean-like soft operators can be applied to combine the input data. Specifically, the objective of this contribution is the proposal of soft fusion operators at pixel level to combine heterogeneous time series of remote sensing images that can aid to improve agricultural and farming tasks such as crop classification and monitoring. The approach that we propose in this contribution is based on the approach in [5] using soft operators. This choice is motivated by the fact that we want to define a feasible and low cost (in terms of time and memory needed to process data) fusion of heterogeneous time series. More precisely we state some desired properties of the synthetized image to be generated and then define soft fusion operators that model these desired behaviors of the fusion. We introduce the fusion operators by starting with simpler ones, and adding assumptions so as to satisfy distinct increasing properties. Finally, we present some results when applying the soft operators to fuse heterogeneous time series of NDVI (*Normalized Difference Vegetation Index*) images relative to two distinct geographic regions (Brazil and Italy) characterized by distinct vegetation cover and climate. We compare the synthesized images with the real images we pretend to predict at given dates with those obtained by the application of a fuzzy rule-based approach [3,4].

## 2 Problem Formulation: Soft Fusion of Time Series of Images

Let us consider two time series of images: the first one <H> consisting of a sparse sequence of high resolution images $H_1$, …, $H_n$ and the second dense series <L> of low resolution images $L_1$, …, $L_m$, with n<<m.

Let us consider that the pixel values $h \in H$ and $l \in L$ are defined for both images in [0,1] and represent some vegetation index such as the NDVI. The NDVI represents the density of green leaves on the surface and takes values between 0 and 1 for bare soil and vegetation and negative values for water.

Further, for the high resolution images $H_1$, $H_2$, …, $H_n$ we know the exact timestamp $t_{H1}$, $t_{H2}$, …, $t_{Hn}$, while this is not the case for the second series of low resolution images which are often built by a composition of several images taken at distinct timestamps within distinct subsequent time intervals $[t_{L1min}, t_{L1max}]$, …$[t_{Lm\ min}, t_{Lm\ max}]$. This is actually a realistic hypothesis when the two series are Landsat images and MODIS (*MODerate Imaging Spectroradiometer*) images respectively.

It is well known that the objectives of image fusion may be very different: they span from image brightness enhancement, to edge enhancement, to objects segmentation and change detection. In the case we are tackling, the objectives can be to generate a denser image time series to better represent the evolution (changes) of some dynamic phenomenon. For instance that could be crop growth or improvement in classification results exploiting more images at specific timestamps. Finally, one can consider the fusion of multiple heterogeneous images from the two time series or just two images, one from the dense and the other from the sparse series.

In the first step, for sake of clarity, we assume to define a fusion function of two input images.

### 2.1 Soft Fusion Considering the Temporal Validity (WA)

Hereafter, we state the objective of the fusion starting from the simplest assumptions and then adding new requirements so as to define step by step fusion functions with increasing complexity.

The objective of the fusion is generating a synthetic high resolution image which is lifelike for any desired timestamp *t*.

The fusion function is defined as F: [0,1] x [0,1] → [0,1]; such that given a pair of unit values *l* and *h* it returns *h'*: *h'=F(l,h)*

The first desired property of the F function is the following:

- The fused image H' must be generated by considering its temporal validity at the desired timestamp *t*. This means that we must find a method to weigh differently the contributions of H and L depending on the temporal distance of their timestamps $t_H$ and $t_L$ from *t* so that :
  - if $|t_H-t| < |t_L-t|$ then H' is mainly determined by H and marginally by L
  - else then H' is mainly determined by L and marginally by H
    where *mainly* and *marginally* are defined as weights that vary smoothly with the variations of $|t_H-t|$ and $|t_L-t|$ .

A simple fusion function satisfying this property is a weighted average in which the weight is based on the temporal information associated with the input images H and L and the desired date *t* of the image we want to generate H'. It is reasonable to assume that the smaller the interval of time between the desired time instant *t* and the timestamps of the input images, the greater could be their contributions. So the contribution can be expressed by a weight that is inversely proportional to $|t-t_H|$ and $|t-t_L|$ for H and L respectively.

To this end, in absence of knowledge on the dynamics of the observed elements in the images we can imagine to define a triangular membership function $\mu_t$ of a fuzzy set on the temporal domain of the time series, with the central vertex in *t* and the other two vertexes $t_O$ and $t_E$ placed outside of the temporal domain $[t_H, t_L]$.

Another possibility is to choose $t_O$ and $t_E$ based on the knowledge of the expert: it can be a time range in which the dynamics of the contents of the images to select for the fusion do not change drastically. For example, in the case of the NDVI one could define $t_O$ and $t_E$ as the time limits of the season to which *t* belongs to. In fact, fusing two NDVI images, one taken in winter and the other in summer time, would be meaningless since the characteristics of some vegetation types may be completely different in these two seasons.

The choice of the membership degree $\mu_t(t_H)$ and $\mu_t(t_L)$ of time stamps $t_H$ and $t_L$, which are normalized in [0,1], can be taken as the weights defining the temporal validities of the signal in H and L at the desired time instant *t* and thus can define the contributions of the corresponding image H and L to the synthetic image H' that we want to compute. The situation is depicted in Figure 1 where at timestamp t we generate a fused image by combining the two images $L_4$ and $H_1$ since these are the closest in time to *t*.

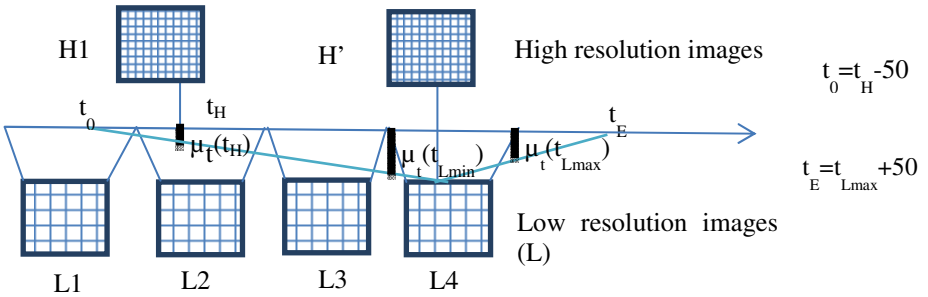



**Fig. 1.** Two time series of High and Low resolution images are depicted, with distinct timestamps and temporal density. In black the membership degrees defining the time validity of the images with respect to timestamp t.

Once one has set the range of the temporal domain in which it makes sense to select the images to fuse: i.e., $t_O$ and $t_E$, the validity of a timestamp $t_H$ can be computed as follows:

- If $t_0 \leq t_H < t$ then $\mu_t(t_H) = (t_H - t_0)/(t - t_0)$
- If $t \leq t_H < t_E$ then $\mu_t(t_H) = (t_H - t_E)/(t - t_E)$ (1)

We can also compute the validity degree of a low resolution image L at an instant of time *t* by computing its membership degrees $\mu_t(t_{Lmin})$ and $\mu_t(t_{Lmax})$ by applying equation (1) and then taking the greatest validity degree, i.e., we take the best validity of the two extremes of the time range of L:

$$\mu_t(t_L)=\max(\mu_t(t_{Lmin}), \mu_t(t_{Lmax})), \quad (2)$$

We then select from the two series the H image and the L image with greatest validities in their series.

A fusion function that satisfies the above stated property is the following:

- Given two time series <H> and <L> and a desired timestamp *t* we first select the input images to fuse H∈<H> and L∈<L> such that:
  $H = \text{argMax}_{H\in<H>}(\mu_t(t_{H1}),\ldots, \mu_t(t_{Hm}))$ and $L = \text{argMax}_{L\in<L>}(\mu_t(t_{L1}),\ldots, \mu_t(t_{Lm}))$

Then, for each pixel values $l \in$ L and $h \in$ H we compute the Weighted Average:

$$h' = WA(h,l) = \frac{f(\mu_t(t_L))*l + f(\mu_t(t_H))*h}{f(\mu_t(t_L)) + f(\mu_t(t_H))} \quad (3)$$

where $f(.) \in [0,1]$. *f* is a function associated with a linguistic modifier such as *very*, *fair*, *more or less, idem* etc, i.e., a concentrator or dilator of its argument. It is generally defined as $f(.)=(.)^x$ with $x>1$, in case of concentrator, while $x<1$ in case of dilator. The choice of *x* must be tuned based on sensitivity analysis exploiting training data, i.e., by comparing the correlation obtained between the resulting image H' and the expected image E at timestamp *t*, a target image that we have for timestamp *t*.

In the case the expert knows the dynamics of represented objects; *f* can be defined by a membership function describing the temporal validity of an image taken in a given period.

Notice that equation (3) can be easily generalized to fuse K input images, provided that their temporal validities are computed by applying equation (1) to their timestamps and that they are selected from the two time series based on the following:

- $H=\text{argMax\_K}_{H\in<H>}(\mu_t(t_{H1}),\ldots,\mu_t(t_{Hm}))$, $L=\text{argMax\_K}_{L\in<L>}(\mu_t(t_{L1}),\ldots,\mu_t(t_{Lm}))$ (4)
  where argMax_K selects the H which has the validity degree $\mu_t(t_H)$ within the K greatest ranked values.

Notice that, fusing more than two images to generate a simulated image can be useful to cope with the very frequent problem of clouds masking the scene in the original images of the time series.

### 2.2 Soft Fusion Based on Temporal Validity and Preference (WP)

Let us assume that we know that one of the two time series is better than the other one with respect to the signal in the images, either because it is less affected by noise or because it has undergone preprocessing steps that have cleansed it.

- In this case, besides the temporal validity, we want to model in the fusion a preference for one of the two series. Assuming that <H> is preferred to <L> we can formalize the desired properties with the following rule:
  - if $\mu_t(t_H)$ is *very high* then H' is determined *mostly* by H and *marginally* by L, else it's a weighted average of H and L with contributions directly proportional to their temporal validities.

    *very high* can be quantified by a positive integer $p$, a numeric preference indicating how many times H is preferred with respect to L. *mostly* and *marginally* can be flexibly tuned by modifying the validity degrees of L and H depending on $p$, so that the validity of L is decreased while the validity of H is increased.

    We also want not to overestimate the pixel values.

A fusion function that satisfies the above stated properties is the following:
given two pixel values $l$ and $h$ for L and H respectively:

$$h' = WP(h,l) = min\left(\begin{matrix} max\left(WA(h,l), \mu_t(t_H)\right), \\ \frac{(\mu_t(t_L))^p * l + (\mu_t(t_H))^{1/p} * h}{(\mu_t(t_L))^p + (\mu_t(t_H))^{1/p}} \end{matrix}\right), p \geq 1 \quad (5)$$

In which $WA(h,l)$ is defined as in equation (3). While *WA* is symmetric *WP* is an asymmetric function. The asymmetry of *WP* function depends on both the fact that p>1, and the satisfaction of the condition $WA(h,l)>\mu_t(t_H)$. We can observe that when $\mu_t(t_H) = 1$, the preferred value $h$ has a greater chance to contribute to $h'$, and its contribution increases with the preference $p$, unless when $p=1$ in which case we get exactly the Weighted Average based solely on temporal validities of H and L: $h'=\min(1, WA(h,l))$ . When $WA(h,l)>\mu_t(t_H)$ and p>1 we get the minimum between the Weighted Average and the Weighted Preferences.

A dual function satisfying the following properties can be defined:

- Assuming that <H> is preferred to <L>:
  - if μt($t_H$) is *very high* then H' is determined *mostly* by H and *marginally* by L, else it's a weighted average of H and L with contributions directly proportional to their temporal validities.
- We also want not to underestimate the pixel values

The corresponding function is the following:

$$h' = WP(l,h) = max\begin{pmatrix} min\big(WA(h,l), 1-\mu_t(t_H)\big), \\ \frac{(\mu_t(t_L))^p * l + (\mu_t(t_H))^{1/p} * h}{(\mu_t(t_L))^p + (\mu_t(t_H))^{1/p}} \end{pmatrix}, p \geq 1 \quad (6)$$

We can observe that when$(\mu_t(t_H))$ tends to 1, the preferred value $h$ has a greater chance to contribute to $h'$ and its contribution increases with $p$. On the contrary we obtain the weighted average based solely on the temporal validities of H and L.

For the application of this method, some conditions have to be defined for choosing between equation 5 and equation 6. In our experiments we wanted to identify if the input images correspond to a growing or senescent season. In the case of a growing season we do not want to underestimate the NDVI values so we can choose equation 6. Contrarily, in the senescent season we do not want to overestimate so equation 5 must be applied. The identification of the season was made automatically from the timestamps of the images and the tendency of the average NDVI in the time interval between the input images. The method was applied as follows:

- Senescent season: If $t_H < t < t_L$ and $\Sigma_{h\in H}\, h/(n*m) > \Sigma_{l\in L}\, l/(n*m)$ or

if $t_L < t < t_H$ and $\Sigma_{l\in L}\, l/(n*m) > \Sigma_{h\in H}\, h/(n*m)$ then apply equation $WP(h,l)$ (5)

- Growing season: If $t_H < t < t_L$ and $\Sigma_{h\in H}\, h/(n*m) < \Sigma_{l\in L}\, l/(n*m)$ or

if $t_L < t < t_H$ and $\Sigma_{l\in L}\, l/(n*m) < \Sigma_{h\in H}\, h/(n*m)$ then apply equation $WP(l,h)$ (6)

where $n*m$ is the total number of pixels . This method could be applied differently by changing the conditions for applying equation 5 or 6 and also by selecting a value of $p$ lower than 1 if we wanted to give preference to the low resolution image.

### 2.3 Soft Fusion Considering the Temporal Validity and the Change (WS)

The previous soft fusions do not consider the change of the pixel values, i.e., $|h-l|$ while this would be desirable when one wants to enhance edges of change. They typically correspond to those regions where there is a high increase or decrease of the pixel values.

The change can be defined by a positive value as follows:

$$s = \min\left[(|h-l| - s_{min}) / (s_{max\,p\%} - s_{min}), s_{max}\right]$$

where $s_{min}$ and $s_{max}$ are the minimum and maximum $|h-l|$ among all pixels in H and L and $s_{max\,\%p}$ is the maximum $|h-l|$ for %p percentile in H and L.

The desired property is the following:

- The more $s$ is high and the more $H$ is valid ( $\mu_t(t_H)$ is high) then the more $h'$ should be close to $h$

A simple soft fusion function satisfying the above properties is the following:

$$h' = WS(h,l) = \frac{(1-s)*\mu_t(t_L)*l + s*\mu_t(t_H)*h}{(1-s)*\mu_t(t_L) + s*\mu_t(t_H h)} \quad (7)$$

# 3 Application of the Methods to Real Data

## 3.1 Data

A short temporal series of high and low resolution images was available for a zone of Brazil for the year 2004. As low resolution images we used the MODIS product MOD13Q1 which provides NDVI composite images each 16 days at 250m spatial resolution. As high resolution images we had CBERS (China-Brazil Earth Resources Satellite) images with 20 m spatial resolution. MODIS images were available at dates 225, 241, 257 and 273 (starting date of the 16-day period). CBERS images were available for dates 228, 254 and 280. The CBERS images were radiometrically normalized to the MODIS ones as described in [6]. Also, several images of a zone of Italy were available for the year 2012. As low resolution images we had one image (date 209) of the MODIS product MOD09Q1 which provides composite images of 8 days at 250m spatial resolution. As high resolution we had two Landsat images (with 30m spatial resolution) in the dates 197 and 213.

## 3.2 Results and Quality Evaluation

The soft operators defined in section 2 were applied to different temporal combinations of the available images.

The simulated images with the different methods (Weighted Average: WA, Weighted with Preference: WP, and Weighted with Slope: WS) were then compared to the 'target' (high resolution image in the date we pretend to simulate) and the following quality indices were computed: correlation, RMSE (Root Mean Square Error) and the accuracy between the simulated and the target image, defined as follows:

$$Accuracy = 1 - \frac{\sum_i |h'_i - t_i|}{n * m}$$

in which $h'_i$ and $t_i$ are the pixel values corresponding to the simulated and the target images respectively, and $n*m$ is the total number of pixels.

The different temporal combinations for analyzing the proposed methods are shown in Table 1. In Table 2 we show the results obtained with the previous methods in the different temporal combinations of the available images. Analyzing the correlations we observe that the fuzzy rule-based method using three rules equally distributed leads to significantly lower correlations than the different proposed methods using soft operators. When analyzing the proposed methods we observe that in the cases where the input and target images are far away (cases of Brazil T5 and T6: a difference of 52 days) the method leading to the higher correlations is the WA. In the other cases with closer dates (Brazil T1, T2 and T3: difference of 26 days) the results are not so clear, low differences are observed between the different methods using soft operators. Only in the case of the input high resolution date being lower than the target (case T4: -26 days) we observe a better correlation when using the method WP with preference $p$=2. Conversely, in Italy's zones the WA leads to lower correlations than the other soft operators, while the WS and WP with $p=2$ lead to the higher correlations. However we keep observing lower values when using the fuzzy rule-based method in Italy Z1. Regarding the RMSE we observe generally higher errors when using the fuzzy rule-based method (6 out of the 8 cases). In the methods using soft operators we observe that in the further images (cases of Brazil T5 and T6) we obtain again the lower errors when using the WA. In the other cases the best

method is difficult to identify, the WA would be clearly selected in cases Brazil T2 and T3. In the other cases there are other methods with similar results to the WA. We observe in 5 out of the 6 cases of Brazil that the WP with *p=2* leads to the higher errors, while in Italy's zone 1 this method together with the WS lead to the lowest values of RMSE. In Italy's zone 2 there are no significant differences between the soft operators. The Italian zone 2 covers a very dense urban area and thus the setting of the time validity is probably not very appropriate for this scene.

**Table 1.** Temporal combinations used for applying the methods. Timestamps are expressed as day of the year. Hr corresponds to the high resolution image used in the fusion algorithm, $Lr_{min}$ and $Lr_{max}$ are the minimum and maximum timestamps of the composite MODIS image, Target is the image to simulate, and Dist(t) is the difference in days between the timestamps of the high resolution input image and the Target image.

| | Italy Z1 | Italy Z2 | Brazil T1 | Brazil T2 | Brazil T3 | Brazil T4 | Brazil T5 | Brazil T6 |
|---|---|---|---|---|---|---|---|---|
| **Hr** | 197 | 197 | 254 | 228 | 228 | 254 | 228 | 280 |
| **$Lr_{min}$** | 209 | 209 | 273 | 241 | 257 | 225 | 273 | 225 |
| **$Lr_{max}$** | 216 | 216 | 288 | 256 | 272 | 240 | 288 | 240 |
| **Target** | 213 | 213 | 280 | 254 | 254 | 228 | 280 | 228 |
| **Dist(t)** | 16 | 16 | 26 | 26 | 26 | -26 | 52 | -52 |

**Table 2.** Results of the soft fusion methods obtained in the different temporal combinations. High values of both R and Accuracy and low values of RMSE as associated with good quality of the simulated image with respect to the target image. The best quality indicators triples (R, RMSE, Accuracy) for each simulated image are reported in bold cases.

| | | Italy Z1 | Italy Z2 | Brazil T1 | Brazil T2 | Brazil T3 | Brazil T4 | Brazil T5 | Brazil T6 |
|---|---|---|---|---|---|---|---|---|---|
| **WA** | R | 0.87 | **0.83** | 0.91 | 0.85 | 0.90 | **0.89** | **0.89** | **0.86** |
| | RMSE | 0.10 | **0.14** | 0.06 | 0.08 | 0.07 | **0.11** | **0.07** | **0.10** |
| | Accuracy | 0.93 | **0.89** | 0.95 | 0.94 | 0.94 | **0.90** | **0.95** | **0.91** |
| **WS** | R | **0.89** | **0.83** | 0.91 | **0.86** | **0.90** | **0.89** | 0.88 | 0.84 |
| | RMSE | **0.09** | **0.14** | 0.06 | **0.09** | **0.08** | **0.11** | 0.07 | 0.11 |
| | Accuracy | **0.93** | **0.89** | 0.95 | **0.94** | **0.94** | **0.90** | 0.95 | 0.91 |
| **WP, p=1.3** | R | **0.88** | **0.83** | 0.91 | **0.86** | **0.90** | **0.89** | **0.89** | 0.86 |
| | RMSE | **0.10** | **0.14** | 0.06 | **0.09** | **0.08** | **0.11** | **0.07** | 0.11 |
| | Accuracy | **0.93** | **0.89** | 0.95 | **0.94** | **0.94** | **0.90** | **0.95** | 0.91 |
| **WP, p=1.5** | R | **0.88** | **0.83** | **0.92** | 0.86 | **0.90** | **0.89** | 0.88 | 0.86 |
| | RMSE | **0.10** | **0.14** | **0.06** | 0.09 | **0.08** | **0.11** | 0.07 | 0.11 |
| | Accuracy | **0.93** | **0.89** | **0.95** | 0.93 | **0.94** | **0.90** | 0.95 | 0.91 |
| **WP, p=1.7** | R | **0.88** | **0.83** | **0.92** | 0.86 | **0.90** | **0.89** | 0.88 | 0.86 |
| | RMSE | **0.10** | **0.14** | **0.06** | 0.09 | **0.08** | **0.11** | 0.07 | 0.11 |
| | Accuracy | **0.93** | **0.89** | **0.95** | 0.93 | **0.94** | **0.90** | 0.95 | 0.91 |
| **WP, p=2** | R | **0.88** | **0.83** | **0.92** | 0.86 | 0.89 | **0.90** | 0.87 | 0.84 |
| | RMSE | **0.10** | **0.14** | **0.06** | 0.09 | 0.08 | **0.12** | 0.08 | 0.11 |
| | Accuracy | **0.93** | **0.89** | **0.95** | 0.93 | 0.93 | **0.90** | 0.94 | 0.90 |
| **Fuzzy** | R | 0.87 | 0.83 | 0.89 | 0.83 | 0.88 | 0.88 | 0.85 | 0.85 |
| | RMSE | 0.11 | 0.14 | 0.07 | 0.10 | 0.09 | 0.11 | 0.09 | 0.10 |
| | Accuracy | 0.91 | 0.88 | 0.95 | 0.92 | 0.93 | 0.91 | 0.93 | 0.92 |

In terms of accuracy, the fuzzy rule-based methods leads to similar results as the RMSE, showing the lower accuracies in 6 out of the 8 cases (the same having higher RMSE values) and the higher accuracies in the other two cases. The WP with *p=2* leads generally to lower accuracies, and the WA leads to high accuracies in the Brazil zone. In Italy's zones 1 we obtain higher accuracies with the WS and the WP with *p=2*, while in zone 2 all the methods lead to similar results. We can conclude that the WA is the one showing the better compromise between correlation, RMSE and accuracy in the different temporal combinations of the Brazil's zone. However if we want to use the fused images for edge detection it is not so important having high accuracy but it is important having a high correlation, so we could use the method WP with *p=2* in the case of close images. In Italy's zone 1 the methods WS and WP with *p=2* are the best ones in terms of correlation, RMSE and accuracy, while in zone 2 all the methods lead to similar results.

## 4 Conclusions

The paper proposes some soft fusion operators to generate synthetized images at desired timestamps having two input heterogeneous time series of remote sensing data. The proposed operators were applied to different combinations of input and target images. A fuzzy rule-based fusion method was also applied to the same combinations of images. The validation of the results obtained with the different operators as well as the comparison with the fuzzy rule-based fusion method were analysed in terms of correlation between the simulated and target images, RMSE and accuracy. The proposed fuzzy operators led to higher correlations than the fuzzy rule-based method applied in all the cases and to higher (lower) values of accuracy (RMSE) in most of the cases. These results show how the application of simple soft operators taking into account the time validity of input images and in some cases a preference for the high resolution input image can be used for simulating images with high accuracy and correlation at desired timestamps within the timestamps of the input images.

**Acknowledgments.** This work was supported by public funds received in the framework of GEOSUD, a project (ANR-10-EQPX-20) of the program 'Investissements d'Avenir' managed by the French National Research Agency, the Italian CNR short term mobility program and the Space4Agri project of CNR and Regione Lombardia.

## References

1. Goshtasby, A.A., Nikolov, S.: Guest editorial: Image fusion: Advances in the state of the art. Inf. Fusion 8, 114–118 (2007)
2. Pohl, C., Van Genderen, J.L.: Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. Int. J. Remote Sens. 19, 823–854 (1998)

3. Dammavalam, S.R.: Quality Assessment of Pixel-Level ImageFusion Using Fuzzy Logic. Int. J. Soft Comput. 3, 11–23 (2012)
4. Zhao, L., Xu, B., Tang, W., Chen, Z.: A Pixel-Level Multisensor Image Fusion Algorithm Based on Fuzzy Logic. In: Wang, L., Jin, Y. (eds.) FSKD 2005. LNCS (LNAI), vol. 3613, pp. 717–720. Springer, Heidelberg (2005)
5. Yager, R.R.: A framework for multi-source data fusion. Soft Comput. Data Min. 163, 175–200 (2004)
6. Le Maire, G., Marsden, C., Verhoef, W., Ponzoni, F.J., Lo Seen, D., Bégué, A., Stape, J.-L., Nouvellon, Y.: Leaf area index estimation with MODIS reflectance time series and model inversion during full rotations of Eucalyptus plantations. Remote Sens. Environ. 115, 586–599 (2011)

# Fuzzy Argumentation System for Decision Support

Nouredine Tamani and Madalina Croitoru

INRIA GraphIK, LIRMM, 161 rue Ada, Montpellier, France
{tamani,croitoru}@lirmm.fr

**Abstract.** We introduce in this paper a quantitative preference based argumentation system relying on ASPIC argumentation framework and fuzzy set theory. The knowledge base is fuzzified to allow the experts to express their expertise (premises and rules) attached with grades of importance in the unit interval. Arguments are attached with a score aggregating the importance expressed on their premises and rules. Extensions are then computed and the strength of each of which can also be obtained based on its strong arguments. The strengths are used to rank fuzzy extensions from the strongest to the weakest one, upon which decisions can be made. The approach is finally used for decision making in a real world application within the EcoBioCap project.

## 1 Introduction

Logically instantiated argumentation system (such as ASPIC/ASPIC+ framework [1, 2]) can be used to reason with arguments to provide means (i) to express argument as a combination of premises and inference rules, (ii) to define contradictions, attacks and defeat between arguments and (iii) to extract consistent subsets of arguments called extensions that also defend themselves against attacks. In practice, the decision-maker has often to deal with several extensions leading to conflicting decisions. In this context the argumentation process is not enough, since it cannot say what is the best extension to consider to make a decision. To address this area, several approaches for defining preference-based argumentation systems have been proposed during the last years [2–4]. These approaches extend the Dung abstract argumentation model by introducing a preorder or a total order on arguments through a preference relation, which states for each couple (or even for two subsets) of arguments either they are incomparable or which is the most preferred. In a logically instantiated argumentation framework ([2–4]), this preference order is used, among other, for a qualitative ranking of extensions. However, in real world application, the user may be unable to specify all the relative importance of arguments, because in the worst case complexity of computation of the ranking is exponential.

***Contribution.*** We introduce in this paper a quantitative approach to define a total order between extensions in logically instantiated argumentation framework. We propose to model the premises and rules as fuzzy sets such that the

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 77–86, 2014.

grade of membership of a premise (resp. a rule) expresses its importance from expert standpoints. The notions of strictness and defeasibility are then unified under the notion of importance. The rules of the system are then more expressive in the sense that the closer to 1 the importance of a premise (resp. a rule) is, the more strict it is, and conversely the closer to 0 a premise (resp. a rule) is, the more defeasible it is. Then, arguments are attached with a strength score based on the degrees of membership of the involved premises and rules, which can be computed with a t-norm. Based on the argument strengths, we can compute a force of each extension $E$, by using the decomposition based approach [5], corresponding to the truth value of the fuzzy quantified proposition "*almost_all arguments in E are strong*". We obtain a total order between extensions used to rank them from the strongest to the weakest one in a polynomial time. The proposed approach is finally applied within the EcoBioCap project to show its usefulness in decision making for packaging selection.

In section 2, we recall the principles of argumentation framework. In section 3, the rationale behind the fuzzy ASPIC argumentation system is introduced. Its application in a real world application is detailed in Section 4. Section 5 summarizes related work in the field and finally Section 6 concludes the paper.

## 2 Argumentation Framework

### 2.1 Dung Argumentation Principles

Dung argumentation framework ($AF$) [6] is a tuple $(\mathcal{A}, \mathcal{C})$, where $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$ is a binary attack relation on the set of arguments $\mathcal{A}$, having a meaning of defeat. For each argument $X \in \mathcal{A}$, $X$ is *acceptable* w.r.t. a set of arguments $S \subseteq \mathcal{A}$ iff any argument attacking $X$, is attacked by an argument of $S$. A set of arguments $S \subseteq \mathcal{A}$ is *conflict free* iff $\forall X, Y \in S, (X, Y) \notin \mathcal{C}$. For any conflict free set of arguments $S$, $S$ is an *admissible extension* iff $X \in S$ implies $X$ is acceptable w.r.t. $S$; $S$ is a *complete extension* iff $X \in S$ whenever $X$ is acceptable w.r.t. $S$; $S$ is a *preferred extension* iff it is a set inclusion maximal complete extension; $S$ is the *grounded extension* iff it is the set inclusion minimal complete extension; $S$ is a *stable extension* iff it is preferred and $\forall Y \notin S, \exists X \in S$ such that $(X, Y) \in \mathcal{C}$.

For $T \in \{$complete, preferred, grounded, stable$\}$, $X$ is skeptically (resp. credulously) justified under the $T$ semantics if $X$ belongs to all (resp. at least one) $T$ extension. Output of an extension $E$ is $Output(E) = \{Conc(A), A \in E\}$, where $Conc(A)$ is the conclusion of argument $A$. The skeptical output of $AF$ is $Output(AF) = \bigcap_{i=1,\ldots,n} Output(E_i)$ such that $E_i$ are its $T$ extensions.

### 2.2 ASPIC *AS* Argumentation System (ASPIC *AS*)

In this paper we consider ASPIC $AS$ as a subset of ASPIC+ [7] and compatible with the one presented in [8]. An ASPIC $AS$ is a tuple $(\mathcal{L}, cf, \mathcal{R}, \geq)$, where $\mathcal{L}$ is a logical language, $cf$ is a contrariness function which associates to each formula of $\mathcal{L}$ a set of its incompatible formulas (in $2^{\mathcal{L}}$), $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is the set of strict

($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) rules of the form $\varphi_1, ..., \varphi_m \rightarrow \varphi$ and $\varphi_1, ..., \varphi_m \Rightarrow \varphi$ respectively, where $\varphi_{i,i=1,...,m}, \varphi$ are well-formed formulas in $\mathcal{L}$, and $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$, and $\geq$ is a preference ordering over defeasible rules.

A knowledge base in $AS = (\mathcal{L}, cf, \mathcal{R}, \geq)$ is $\mathcal{K} \subseteq \mathcal{L}$ such that $\mathcal{K} = \mathcal{K}_a \cup \mathcal{K}_p$ and $\mathcal{K}_a \cap \mathcal{K}_p = \emptyset$, $\mathcal{K}_a$ contains axioms and $\mathcal{K}_p$ contains ordinary premises.

## 3 Fuzzy ASPIC Argumentation System (F-ASPIC AS)

### 3.1 Fuzzy Set Theory

Fuzzy set theory introduced by Zadeh [9] to express the gradual membership of an element to a set. Formally, a fuzzy set $F$ is defined on a referential $U$ by a membership function $\mu_F : U \mapsto [0, 1]$ such that $\mu_F(x)$ denotes the membership grade of $x$ in $F$. In particular, $\mu_F(x) = 1$ denotes the full membership of $x$ in $F$, $\mu_F(x) = 0$ expresses the absolute non-membership and when $0 < \mu_F(x) < 1$, it reflects a partial membership (the closer to 1 $\mu_F(x)$, the more $x$ belongs to $F$). The core of $F$ is $Core(F) = \{x \in F : \mu_F(x) = 1\}$ and its support is $Support(F) = \{x \in F : \mu_F(x) > 0\}$. If a fuzzy set is a discrete set then it is denoted $F = \{(x_1, \mu_F(x_1)), ..., (x_n, \mu_F(x_n))\}$, otherwise, it is characterized by its membership function, in practice often a trapezoidal function. The union $\cup$ and the intersection $\cap$ operators are defined by a couple of a t-norm and a t-conorm, such as (min, max). Let $F$, $G$ be two fuzzy sets, $\mu_{F \cup G}(x) = max(\mu_F(x), \mu_G(x))$, $\mu_{F \cap G}(x) = min(\mu_F(x), \mu_G(x))$, and the complement of $F$, denoted $F^c$, is $\mu_{F^c}(x) = 1 - \mu_F(x)$. The logical counterparts of $\cap$, $\cup$ and the complement are respectively $\wedge, \vee$ and $\neg$. Other operators have also been defined such as fuzzy implications [10].

### 3.2 Fuzzy Argumentation System: F-ASPIC

A fuzzy argumentation system $FAS$ is equipped with a fuzzy membership function *imp* expressing for each premise and rule its importance. It is worth noticing that the rules do not model fuzzy implications but regular material implications attached with a score expressing a preference order between them reflecting the importance given by the experts.

A fuzzy ASPIC argumentation system is a $FAS = (\mathcal{L}, cf, \mathcal{R}, imp)$ such that $\mathcal{L}$ is a logical language, $cf$ is a contrariness function (we consider the negation $\neg$ as its basic form), $\mathcal{R}$ is the fuzzy set of *important* rules of the form $(\phi_1, ..., \phi_m \rightarrow \phi, s)$ with $\phi_{i,i=1,...,m} \in \mathcal{L}$ are the premises of the rule, $\phi \in \mathcal{L}$ is its conclusion and $s \in [0, 1]$ is its importance, provided by the experts of the domain. For a given rule $r$, if $\mu_{imp}(r) = 1$ then $r$ is a strict rule, if $\mu_{imp}(r) = 0$ then $r$ is an insignificant rule, discarded by the system. If $\mu_{imp}(r) \in ]0, 1[$ then the closer to 1 $\mu_{imp}(r)$ is, the more important $r$ is, and conversely the closer to 0 $\mu_{imp}(r)$ is, the more defeasible $r$ is.

In the same way, a knowledge base $\mathcal{K}$ in a $FAS$ is a fuzzy set of *important* premises. For a given premise $p$, if $\mu_{imp}(p) = 1$ then it is an axiom, if $\mu_{imp}(p) = 0$ then $p$ is insignificant (then discarded by the system), and If $\mu_{imp}(p) \in ]0, 1[$ then the closer to 1 $\mu_{imp}(p)$ is the more important $p$ is.

**F-ASPIC Arguments.** An F-ASPIC argument $A$ can be of the following forms:

1. $\emptyset \rightsquigarrow^s c$, with $c \in \mathcal{K}$, $Prem(A) = \{c\}$, $Conc(A) = c$, $Sub(A) = \{c\}$, $Rules(A) = \emptyset$, where $Prem$ returns premises of $A$, $Conc$ returns its conclusion, $Sub$ returns its the sub-arguments, and $Rules$ returns the rules involved in $A$, $s \in ]0, 1]$ expresses the strength of $A$ (defined below),
2. $A_1, ..., A_m \rightsquigarrow^s c$, such that there exists a rule $r \in \mathcal{R}$ of the form $(Conc(A_1), ..., Conc(A_m) \rightarrow c, s_r)$, and $Prem(A) = Prem(A_1) \cup ... \cup Prem(A_m)$, $Conc(A) = c$, $Sub(A) = Sub(A_1) \cup ... \cup Sub(A_m) \cup \{A\}$, $Rules(A) = Rules(A_1) \cup ... \cup Rules(A_m) \cup \{r\}$, $s \in ]0, 1]$ expresses the strength of $A$ (defined below).

*Remark 1.* An ASPIC argument has a nested form. A sub-argument is also an argument. To improve the readability, by abuse of notation, we associate to each argument a label made of a capital letter followed by a subscript number. The labels are then used in an argument to refer to its sub-arguments. In this notation, a label followed by colon is not a part of the argument. We make use of the same notation in F-ASPIC.

*Example 1.* Let $AS$ be an ASPIC argumentation system defining the rules $\mathcal{R}_s = \{a, b \rightarrow c\}$ and the ordinary premises $\mathcal{K}_p = \{a, b\}$. The following are arguments in $AS$: • $A_1 : \emptyset \Rightarrow a$ • $A_2 : \emptyset \Rightarrow b$ • $A_3 : A_1, A_2 \rightarrow c$.

**Definition 1 (Strength of an Argument).** *The strength of argument A, denoted $str(A)$ computed efficiently as follows:*

$$str(A) = \begin{cases} \mu_{imp}(c), c \in \mathcal{K} & \text{if } A \text{ is of the form } \emptyset \rightsquigarrow c, \\ \mathbb{T}_{e \in Rules(A) \cup Prem(A)}(\mu_{imp}(e)) & \text{if } A \text{ is of the form } A_1, ..., A_m \rightsquigarrow c, \end{cases}$$

*where $\mathbb{T}$ is a triangular norm such as $min, *$, etc.*

**Definition 2 (Strict/Defeasible Argument).** *An argument $A$ is said strict iff $str(A) = 1$. Otherwise, it is called defeasible.*

**Definition 3 (Consistency of $\mathcal{R}$).** *Let $\mathcal{A}$ be the set of arguments of an F-ASPIC $AS$, the set of rule $\mathcal{R}$ is said consistent iff $\nexists A, B \in \mathcal{A}$, such that $str(A) = str(B) = 1$ and $Conc(A) = \neg Conc(B)$.*

### 3.3 Attacks and Defeat between Arguments

We define the fuzzy rebut and fuzzy undercut attacks, denoted F-rebut and F-undercut respectively.

**Definition 4 (F-Rebut Attack).** *An argument $A$ rebuts argument $B$ if $\exists A_i \in Sub(A)$ and $\exists B_j \in Sub(B)$ such that (i) $Conc(A_i) = \neg Conc(B_j)$, (ii) $str(B_j) < 1$ and (iii) $str(A_i) \geq str(B_j)$.*

**Definition 5 (F-undercut Attack).** *Argument A* undercuts *argument B iff* $\exists B' \in Sub(B)$ *of the form* $B'_1, ..., B'_m \rightsquigarrow^s \varphi$, *with* $s < 1$ *and* $\exists A' \in Sub(A)$ : $Conc(A') = \neg \lceil Conc(B'_1), ..., Conc(B'_m) \rightsquigarrow^s \varphi \rceil$, *with* $str(A') \geq s$, *and operator* $\lceil . \rceil$ *converts a defeasible rule into a literal.*

**Definition 6 (F-Defeat).** *Argument A defeats arguments B iff A F-rebuts or F-undercuts B.*

### 3.4 F-ASPIC AS Extensions and Rationality Postulates

Rationality postulates, defined in [11], ensure the completeness and the consistency of the output of a logical-based argumentation system. We consider here the following rationality postulates:

1. *Closure under sub-arguments*: for every argument in an extension, also all its sub-arguments are in the extension,
2. *Closure under strict rules* $\mathcal{R}_s = \{r \in \mathcal{R} | \mu_{imp}(r) = 1\}$ of the output of an extension: all possible conclusions from applicable strict rules are derived in each extension,
3. *Direct consistency*: the output of each extension is consistent, so it is not allowed to derive a conclusion and its contradiction in an extension,
4. *Indirect consistency*: the closure under strict rules of the output of each extension is consistent.

**Proposition 1.** *F-ASPIC AS is closed under sub-arguments and under strict rules, direct and indirect consistent, iff the set of strict rules* $\mathcal{R}_s$ *is closed under transposition [11] and* $\mathcal{R}_s$ *is consistent.*

The proof of proposition 1 is detailed in the technical report [12] (pages 5-7). Proposition 1 shows that extensions can be built upon our defeat relation without falling in the inconsistency described in [3, 4, 2] when preferences are involved in the defeat relation and not in the attacks. So, our logical-based argumentation system is then compatible with Dung abstract model for argumentation.

### 3.5 Ordering Extensions

Several aggregating operators can be used (*min*, *max*, weighted mean, etc.) to compute the force of an extension combining the strength of its arguments. Although they are simple to implement they can lead to non-justifiable results, as in the case of *max* operator, for instance, an extension containing only strict arguments gets the same force as an extension having only one strict argument and numerous very weak arguments. A similar result can be obtained from a weighted mean operator. So, we need a hybrid operator able to deliver a force taking into account the strength of arguments and their number. Therefore, we define a force of an extension based on fuzzy quantified proposition as follows.

**Definition 7 (Force of an Extension).** *The force of an extension E (denoted* $force(E)$*), under one Dung's semantics, is the truth value of the fuzzy quantified proposition "P :* almost_all arguments in $E$ are strong*", denoted* $\delta_P$.

$\delta_P$ can be computed efficiently in a polynomial time by the decomposition based approach [5]. So arguments in $E$ are ranked from the most to the least strong: $str(A_1) \geq str(A_2) \geq ... \geq str(A_n)$ with $n = |E|$, and

$$\delta_P = max_{i=1,...,n}(min(str(A_i), \mu_{almost\_all}(\frac{i}{n})) \quad (1)$$

**Definition 8 (Ordering Relation between Extensions).** *Let $E_1, ..., E_2$ be extensions, under one Dung semantics, of a F-ASPIC AS. The extension $E_i$ is preferred than the extension $E_j$, denoted $E_i \succeq_e E_j$ iff $force(E_i) \geq force(E_j)$. Its strict counterpart $\succ_e$ is: $E_i \succ_e E_j$ iff $force(E_i) > force(E_j)$.*

## 4 Application to Packaging Selection

### 4.1 Selection of Packaging According the Aspect *End of Life*

The following text arguments about the *end of life* of packagings have been collected during an interview with experts in the domain of waste management:

1. Packaging materials, which are biodegradable, compostable, or recyclable having a low environmental impact are preferred,
2. Life Cycle Analysis (LCA) results are not in favor of biodegradable and compostable materials,
3. Consumers are in favor of biodegradable material because they help to protect the environment,
4. Biodegradable materials could encourage people to throw their packaging in nature, causing visual pollution.

We model these arguments by a simple propositional language as follows:

- $BP$, $CP$ and $RP$ are symbols referring respectively to biodegradable, compostable and recyclable packagings,
- $PEV$, $VP$, $HIP$, $LIP$ are symbols referring to packagings which respectively protect the environment, can cause visual pollution problem, have a high or low environmental impact (according to LCA),
- $ACC$, $REJ$ are symbols referring to the decisions (accepted, rejected).

The fuzzy set of *important* rules contains the following weighted implications (membership degrees are provided by the domain experts):
$\mathcal{R} = \{(BP \rightarrow HIP, 0.9), (CP \rightarrow HIP, 0.9), (RP \rightarrow LIP, 0.9),$
$(BP \rightarrow PEV, 0.4), (BP \rightarrow VP, 0.8), (CP \rightarrow VP, 0.8), (HIP \rightarrow REJ, 0.9),$
$(HIP \rightarrow ACC, 0.1), (LIP \rightarrow ACC, 0.9), (LIP \rightarrow REJ, 0.1),$
$(VP \rightarrow REJ, 0.8), (VP \rightarrow ACC, 0.2), (PEV \rightarrow ACC, 0.8), (PEV \rightarrow REJ, 0.2)\}$.

The premises of the system are defined by the fuzzy set corresponding to the *important* packaging choices $\mathcal{K}_p = \{(BP, 0.9), (CP, 0.9), (RP, 0.9)\}$, having the equal importance 0.9. The arguments derived from $\mathcal{K}_p$ and $\mathcal{R}$ are the following (the strength of arguments are computed by the t-norm $min$):

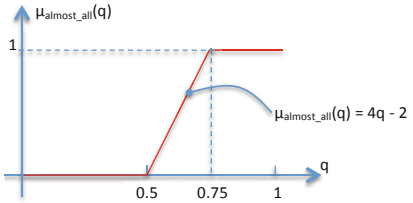



**Fig. 1.** Example of definition of the quantifier *almost_all*

$A_0 : \emptyset \rightsquigarrow^{0.9} BP$
$B_0 : \emptyset \rightsquigarrow^{0.9} CP$
$C_0 : \emptyset \rightsquigarrow^{0.9} RP$
$A_1 : A_0 \rightsquigarrow^{0.9} HIP$
$A_2 : A_1 \rightsquigarrow^{0.9} REJ$
$A_3 : A_1 \rightsquigarrow^{0.1} ACC$
$A_4 : A_0 \rightsquigarrow^{0.4} PEV$
$A_5 : A_4 \rightsquigarrow^{0.4} ACC$
$A_6 : A_4 \rightsquigarrow^{0.2} REJ$
$A_7 : A_0 \rightsquigarrow^{0.8} VP$
$A_8 : A_7 \rightsquigarrow^{0.8} REJ$
$A_9 : A_7 \rightsquigarrow^{0.2} ACC$
$B_1 : B_0 \rightsquigarrow^{0.9} HIP$
$B_2 : B_1 \rightsquigarrow^{0.9} REJ$
$B_3 : B_1 \rightsquigarrow^{0.1} ACC$
$B_4 : B_0 \rightsquigarrow^{0.8} VP$
$B_5 : B_4 \rightsquigarrow^{0.8} REJ$
$B_6 : B_4 \rightsquigarrow^{0.2} ACC$
$C_1 : C_0 \rightsquigarrow^{0.9} LIP$
$C_2 : C_1 \rightsquigarrow^{0.9} ACC$
$C_3 : C_1 \rightsquigarrow^{0.1} REJ$

We can compute the following two fuzzy preferred and stable extensions:

- $\mathcal{E}_1 = \{(A_2, 0.9), (A_8, 0.8), (A_6, 0.2), (B_5, 0.8), (B_2, 0.9), (C_3, 0.1), (B_0, 0.9), (B_1, 0.9), (A_0, 0.9), (A_1, 0.9), (A_4, 0.4), (A_7, 0.8), (B_4, 0.8), (C_0, 0.9), (C_1, 0.9)\}$, and `Output(`$\mathcal{E}_1$`)=`{`REJ,CP,BP,RP,LIP,PEV,HIP,VP`}.
- $\mathcal{E}_2 = \{(A_3, 0.8), (A_5, 0.4), (B_6, 0.2), (C_2, 0.9), (B_3, 0.1), (B_0, 0.9), (B_1, 0.9), (A_0, 0.9), (A_1, 0.9), (A_4, 0.4), (A_7, 0.8), (B_4, 0.8), (C_0, 0.9), (C_1, 0.9)\}$, and `Output(`$\mathcal{E}_2$`)=`{`ACC,CP,BP,RP,LIP,PEV,HIP,VP`}.

Based on formula (1) and subjective definition of the fuzzy quantifier *almost_all* (Figure 1), the force of each extension is $\delta_{\mathcal{E}_1} = 0.8$ and $\delta_{\mathcal{E}_2} = 0.57$, which means that arguments for rejection are stronger than arguments for acceptance. We can here make a decision but it is not justifiable because the rejection of the recyclable packagings is counter-intuitive, since they are supported by strong pros arguments and no arguments are against them.

Independently of the aggregation operator used to compute the force of an extension, the strength of arguments against one type of packaging has an indirect influence on the rejection of the others. To fix the system, we split arguments according to the type of packaging on which they are expressed. So, arguments $A_{i,i=0,\ldots,9}$, $B_{j,j=0,\ldots,6}$ and $C_{k,k=0,\ldots,3}$ are about the acceptance or the rejection of respectively biodegradable (denoted *Bio*), compostable (denoted *Com*) and recyclable (denoted *Rec*) packagings. Then, we draw three argumentation graphs on which we compute preferred and stable extensions, and we obtain:

- $\mathcal{E}_1^{Bio} = \{(A_0, 0.9), (A_1, 0.9), (A_2, 0.9), (A_7, 0.8), (A_8, 0.8), (A_4, 0.4), (A_6, 0.2)\}$, then `Output(`$\mathcal{E}_1^{Bio}$`)= {REJ, BP, HIP, PEV, VP}` resulting to the rejection of biodegradable packagings,
- $\mathcal{E}_1^{Com} = \{(B_0, 0.9), (B_1, 0.9), (B_2, 0.9), (B_4, 0.8), (B_5, 0.8)\}$, so `Output(`$\mathcal{E}_1^{Com}$`)= {REJ, CP, HIP, VP}` leading to the rejection of compostable packagings,

• $\mathcal{E}_1^{Rec} = \{(C_0, 0.9), (C_1, 0.9), (C_2, 0.9)\}$, and `Output(`$\mathcal{E}_1^{Rec}$`)={ACC, RP, LIP}`, supporting the acceptance of recyclable packagings.

This way to deal with arguments allows to make decisions about each packaging choice based on the force of the delivered extensions. So, $\delta_{\mathcal{E}_1^{Bio}} = 0.5$ (average rejection), $\delta_{\mathcal{E}_1^{Com}} = 0.8$ (strong rejection) and $\delta_{\mathcal{E}_1^{Rec}} = 0.9$ (strong acceptance). It is worth noticing that in classical argumentation approach, the user gets the same extensions but no means to compare between them.

**Using the Delivered Extensions in the Querying Process.** From the obtained extensions, we know the reasons why recyclable packagings are recommended and why biodegradable and compostable packagings are not. This result can be used in the querying process by adding to the query the predicate *"recycle = true"* to retrieve the recyclable packaging materials from the database.

## 5 Related work

**Comparison with Preference-Based ASPIC+ [2].** The approach of [2] relies on a preference-based Structured Argumentation Framework (SAF) built upon ASPIC+. An argumentation theory $AT = (AS, \mathcal{K})$ is then a triple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ where $\mathcal{A}$ is the set of all finite arguments of $AS$, $\preceq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff $X$ attacks $Y$. The instantiation of this SAF needs the definition of the operator $\preceq$, which is tricky from the application point of view, since $\preceq$ must satisfy several conditions to be called reasonable [2]. Moreover, ASPIC+ requires the definition of two negations (the contrary and the contradiction), which is not intuitive for the experts providing the arguments.

**Comparison with Fuzzy Argumentation Systems.**Several fuzzy-based argumentation approaches have been proposed during the last decade. In [14–16] possibilistic defeasible logic programming [17] approaches have been developed. They suffer from the difficulties related to the definition of the necessity and possibility measures, since each argument is attached with a necessity degree, and possibilities are attached to the different models of the logical language. In the same context, a possibilistic label-based argumentation approach has been also proposed in [18, 19], expressing the reliability of argument sources. Then, arguments of a single source have a same reliability degree. In [20], a fuzzy argumentation system for trust has been proposed. The approach is similar to ours in the spirit but the context is quite different since in this approach arguments are not structured.

In [21, 22] a fuzzy attack relation is defined to express the strength of attacks between arguments. This approach is also different than ours since we make use of crisp attacks. In [21] the strength of attacks depends on both fuzzy set of arguments supporting the attacker and the strength of the attack. In [22], attack and defeat relations involve a binary preference relation over arguments, in such a way that non-symmetric attacks can fail, resulting to non-intuitive extensions. [23] introduces an argumentation approach based on fuzzy description logic (fuzzy $\mathcal{SHIF}$ DL). Arguments are a combination of fuzzy linguistic

variables and ontological knowledge, and involved in fuzzy attack and support relations. A preference relation is also used. This approach requires users to manually specify the attack and support relations between arguments.

## 6 Conclusion

We have introduced a quantitative preference-based argumentation system relying on the fuzzy sets theory and ASPIC structured argumentation framework. Arguments are built upon the fuzzy set of *important* rules and premises, allowing the computation of their strength. Extensions are attached with a score in $]0, 1]$ aggregating the strength of its contained arguments, based on fuzzy quantified propositions. Finally, the approach is applied on real world application for managing the *end of life* of packaging.

As future work, we plan to implement the approach to study the behavior of the system (i) according to the definition of the linguistic quantifier, (ii) the t-norm used for computing the arguments' strength, and (iii) the variation of the grades of importance associated to premises and rules. In the same line, studying the behavior of the system under a lattice of fuzzy values is worthwhile.

## References

1. Amgoud, L., Bonnefon, J.F., Prade, H.: An argumentation-based approach to multiple criteria decision. In: Godo, L. (ed.) ECSQARU 2005. LNCS (LNAI), vol. 3571, pp. 269–280. Springer, Heidelberg (2005)
2. Modgil, S., Prakken, H.: A general account of argumentation with preferences. Artificial Intelligence 195, 361–397 (2013)
3. Amgoud, L., Vesic, S.: Repairing preference-based argumentation frameworks. In: 21st International Joint Conference on Artificial Intelligence, pp. 665–670 (2009)
4. Amgoud, L., Vesic, S.: Handling inconsistency with preference-based argumentation. In: Deshpande, A., Hunter, A. (eds.) SUM 2010. LNCS, vol. 6379, pp. 56–69. Springer, Heidelberg (2010)
5. Yager, R.R.: General multiple-objective decision functions and linguistically quantified statements. Inter. Jour. of Man-Machine Studies 21(5), 389–400 (1984)
6. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-persons games. Artificial Intelligence 77(2), 321–357 (1995)
7. Prakken, H.: An abstract framework for argumentation with structured arguments. Argument and Computation 1(2), 93–124 (2011)
8. Amgoud, L., Bodenstaff, L., Caminada, M., McBurney, P., Parsons, S., Prakken, H., Veenen, J., Vreeswijk, G.: Final review and report on formal argumentation system.deliverable d2.6 aspic. Technical report (2006)
9. Zadeh, L.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)
10. Bouchon-Meunier, B., Dubois, D., Godo, L., Prade, H.: Fuzzy set and possibility theory in approximate and plausible reasoning. The Handbook of fuzzy sets. In: Fuzzy sets in Approximate Reasoning and Information Systems, pp. 27–31. Kluwer Academic Publishers (1999)

11. Caminada, M., Amgoud, L.: On the evaluation of argumentation formalisms. Artificial Intelligence 171, 286–310 (2007)
12. Tamani, N., Croitoru, M.: Fuzzy argumentation system for decision making. Technical report, INRIA LIRMM (2013), https://drive.google.com/file/d/0B0DPgJDRNwbLdE5wdzFQekJocXM/edit?usp=sharing
13. Destercke, S., Buche, P., Guillard, V.: A flexible bipolar querying approach with imprecise data and guaranteed results. Fuzzy Sets and Systems 169, 51–64 (2011)
14. Chesñevar, C.I., Simari, G.R., Alsinet, T., Godo, L.: A logic programming framework for possibilistic argumentation with vague knowledge. In: Proc. of the 20th Conference on Uncertainty in Artificial Intelligence, UAI 2004, pp. 76–84 (2004)
15. Alsinet, T., Chesñevar, C.I., Godo, L., Simari, G.R.: A logic programming framework for possibilistic argumentation: Formalization and logical properties. Fuzzy Sets and Systems 159(10), 1208–1228 (2008)
16. Alsinet, T., Chesñevar, C.I., Godo, L., Sandri, S., Simari, G.R.: Formalizing argumentative reasoning in a possibilistic logic programming setting with fuzzy unification. International Journal of Approximate Reasoning 48(3), 711–729 (2008)
17. García, A.J., Simari, G.R.: Defeasible logic programming: an argumentative approach. Theory Pract. Log. Program. 4(2), 95–138 (2004)
18. da Costa Pereira, C., Tettamanzi, A.G.B., Villata, S.: Changing one's mind: Erase or rewind? possibilistic belief revision with fuzzy argumentation based on trust. In: Proc. of the 22nd IJCAI, pp. 164–171 (2011)
19. Gratie, C., Florea, A.M.: Fuzzy labeling for argumentation frameworks. In: McBurney, P., Parsons, S., Rahwan, I. (eds.) ArgMAS 2011. LNCS, vol. 7543, pp. 1–8. Springer, Heidelberg (2012)
20. Stranders, R., de Weerdt, M., Witteveen, C.: Fuzzy argumentation for trust. In: Sadri, F., Satoh, K. (eds.) CLIMA VIII 2007. LNCS (LNAI), vol. 5056, pp. 214–230. Springer, Heidelberg (2008)
21. Janssen, J., Vermeir, D., De Cock, M.: Fuzzy argumentation frameworks. In: Proc. of 12th IPMU, pp. 513–520 (2008)
22. Kaci, S., Labreuche, C.: Argumentation framework with fuzzy preference relations. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) IPMU 2010. LNCS, vol. 6178, pp. 554–563. Springer, Heidelberg (2010)
23. Letia, I.A., Groza, A.: Towards pragmatic argumentative agents within a fuzzy description logic framework. In: McBurney, P., Rahwan, I., Parsons, S. (eds.) ArgMAS 2010. LNCS, vol. 6614, pp. 209–227. Springer, Heidelberg (2011)

# Applying a Fuzzy Decision Tree Approach to Soil Classification

Mariana V. Ribeiro[1], Luiz Manoel S. Cunha[2], Heloisa A. Camargo[1], and Luiz Henrique A. Rodrigues[2]

[1] Computing Department of the Federal University of São Carlos
São Carlos, São Paulo, Brazil
{mariana.ribeiro,heloisa}@dc.ufscar.br
http://cig.dc.ufscar.br

[2] College of Agricultural Engineering of the State University of Campinas
Campinas, São Paulo, Brazil
lique@feagri.unicamp.br, luiz.cunha@embrapa.br
http://www.feagri.unicamp.br

**Abstract.** As one of the most important factors that interfere in peoples life, the soil is characterized by quantitative and qualitative features which describe not only the soil itself, but also the environment, the weather and the vegetation around it. Different types of soil can be identified by means of these features. A good soil classification is very important to get a better use of the soil. Soil classification, when performed manually by experts, is not a simple task, as long as the experts opinions may vary considerably. Besides, different types of soil cannot be defined deterministically. With the objective of exploring an alternative approach towards solving this problem, we investigated in this paper the application of an automatic procedure to generate a soil classifier from data, using a fuzzy decision tree induction algorithm. In order to compare the results obtained by means of the fuzzy decision tree classifier, we used two well known methods for classifiers generation: the classic decision tree induction algorithm C4.5 and the fuzzy rules induction algorithm named FURIA.

**Keywords:** fuzzy rule based systems, decision tree, fuzzy decision tree, classification, soil classification, soil classification system.

## 1 Introduction

Due to its use to food cultivation, the soil is one of the most important factors that interfere in people's life, since good food requires good soil. To take advantage of all it's best qualities, not only in food branch, it is very important to know the characteristics of the soil present in each site [20]. Motivated by this, some different classes of soils have been created according to their characteristics. The soil characteristics are related to quantitative and qualitative features that describe the soil, the environment, the weather and the vegetation around them. By knowing its main characteristics, the class that the soil belongs to is

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 87–96, 2014.

also known and then, it is possible to make the best use of it. Although this classification is very useful, it involves very subjective criteria and as it is usually done by experts, it depends very much on the experts opinion. Aiming to support the experts task and reduce the influence of subjectiveness in the classification process, some classification systems can be constructed automatically from data.

Fuzzy decision trees combine the advantages of decision trees, such as the embedded feature selection and low computational cost, with the ability of processing uncertainty and imprecision of fuzzy systems. Some fuzzy decision trees algorithms have been proposed in the literature [2–7]. In this work, we use the fuzzy decision tree induction algorithm (FuzzyDT) described in [2], which is an algorithm based on the well known C4.5 algorithm, to generate fuzzy rules. FuzzyDT starts with the fuzzyfication of the continuous features before inducing the fuzzy decision tree. This algorithm has shown good results in a previous work, when it was applied to a real-world problem, the prediction and control of the coffee rust disease in Brazilian crops [8]. In the work presented here, we investigate the generation of a classification system to deal with the problem of soil classification using FuzzyDT. We also compare the results with the ones obtained by the classic C4.5 algorithm [9] and FURIA algorithm, proposed in [10]. We evaluated them by comparing their accuracy, measured by the correct classification rate, and interpretability, measured by the format and the number of rules generated by each algorithm.

The paper is organized as follows: in the next section, we describe briefly the soil classification problem. In section 3, a short description of decision trees and the main concepts of C4.5 are presented and we describe a general view of the fuzzy classification systems and of the FuzzyDT and FURIA algorithms. The experiments and analyses are presented in section 4. The final conclusions are discussed in section 5.

## 2 Soil Classification

The soil is very important to the human beings and was defined in [1] as "a collection of solid, liquid and gas parts, which could be three-dimensional, moving, formed by minerals and organic materials that occupy most of the surface of the continental extensions of our planet". In Brazil, the soil classification is governed by the Brazilian System of Soil Classification (SiBCS) [1], a hierarchical and multi categorical system, which is open to improvements and expansions.

According to Oliveira [11], researcher and member of the Executive Committee of the Brazilian System of Soil Classification, classifying the soil is very important because it allows:

a) to understand the relation between the individuals
b) to recall the properties of the classified objects
c) to predict the individuals' behavior
d) to improve the use of the soil in a place

e) to estimate the productivity of a stand
f) to provide research themes
g) to explore data from research or observations
h) to facilitate the communication

Actually, the soil classification is extremely important when used to identify the occurrence of different soil in the environment, as in the soil maps.

Nowadays, the SiBCS is constituted by six categorical levels:

1) Order
2) Suborder
3) Large group
4) Subgroup
5) Family
6) Series

So far, the 5th and 6th levels are not organized yet. The attributes which were used in the organization of each level are soils characteristics identified in the research or inferred from other attributes or previous knowledge from soil science. In each categorical level, a set of classes is defined by one or more rules. In this work, we approach specifically the classes Brown Latosol and Brown Nitosol. Brown (redish yellow colors) is a suborder of Latosol and Nitosol orders. Evaluating the soil as Brown Latosol and Brown Nitosol is a crucial problem to the research community because, with the development of the soil science, the understanding of the main diagnostic attributes is under discussion. Diagnostic attributes are characteristics or properties that are used to divide the soil by classification system's levels. Some issues arise in cases where it is difficult to distinguish the soil's characteristics or when it presents conceptual overlap, which hampers the characterization, separation and classification of the soils [12]. The suborder of brown soils has some peculiarities, which demands new investigations that provide a better differentiation among them.

The soil classification task performed by experts started with pedological studies, a practical activity, where over a hundred characteristics' data were collected. These characteristics are defined by quantitative and qualitative data which describe the soil, the environment, the weather and the vegetation around and are used to soil classification. Furthermore, some data were obtained from laboratory analyses done on the collected data, and some other derived from the previous ones.These features were all added to the database in order to complete the set of features that will be used in the classification. Then, these data are discussed by experts which classify the samples based on a predefined pattern of each class and the current soil classification system.

## 3 Classic and Fuzzy Classification Systems

Nowadays, it is very common to deal with a lot of data which are often available on open sources. However, analyzing these data and extracting useful information from them is not an easy task for humans.

In order to solve this problem, some methods of inductive learning have been developed. Among the most used inductive learning methods are the algorithms that generate classifiers. They consist in, given a set of examples, each one described by a set of attributes and a class (or label), learning from these examples and representing the extracted knowledge in a model that is capable of classifying new examples of unknown classes.

A fuzzy system is a system that includes at least one linguistic variable, whose values are linguistic terms represented by fuzzy sets [18].

A very popular and useful type of fuzzy systems are the rule-based fuzzy systems (RBFS), which have a knowledge base, formed by a fuzzy data base and a rule base and an inference mechanism, which processes the rules in the rule base using a reasoning method.

Generally speaking, a classification problem is the problem of assigning a given input data to one of a set of pre-determined set of classes. Rule-based fuzzy classification systems (RBFCS) are a type of RBFS which deals with fuzzy classification problems. After the rules have been constructed, they can be used to classify new instances by applying an inference mechanism such as the ones proposed in [17]. The rules of a RBFCS with n attributes and m classes have the form:

$$\textbf{IF } X_1 \textbf{ is } A_1 \textbf{ AND } X_2 \textbf{ is } A_2 \ldots \textbf{ AND } X_n \textbf{ is } A_n \textbf{ THEN Class is } C_j$$

Where $X_i$ represents the attributes of the set of examples, $A_i$ are the attribute values represented by linguistic terms and $C_j$ is one of the classes in the set of classes $\{C_1, C_2, \ldots C_m\}$.

In the following we describe briefly the learning algorithms used in this work, namely the classic C4.5 algorithm, the FuzzyDT algorithm and the FURIA algorithm.

### 3.1 C4.5 Algorithm

C4.5 is one of the most popular algorithms of decision trees induction. It was proposed by Quinlan [9] and uses entropy and information gain measures to find the most informative attributes for each new split.

The information gain of an attribute is defined as the information that is provided to classification by splitting a set of examples, based on that attribute. It corresponds to its entropy reduction. Higher information gains implies more homogeneous subsets in term of class after splitting. According to Shannon [19], the entropy of a set $S$ containing $k$ possible classes is defined as:

$$E(S) = -\sum_{j=1}^{k} \frac{freq\,(C_j, S)}{|S|} \cdot log_2 \left( \frac{freq\,(C_j, S)}{|S|} \right)$$

Where $freq(C_j, S)$ represents the number of examples in $S$ that belongs to class $C_j$ and $|S|$ is the number of examples in $S$.

The entropy shows the average amount of information necessary to classify an example in $S$.

After splitting $S$ into $n$ subsets $S_i (i = 1, ...n)$ by a node test with attribute $X$ (which is the attribute that provided the highest information gain) the information gain InfGain is given by S's entropy reduction [19]:

$$InfGain(X) = E(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} E(S_i)$$

Once the tree is induced, C4.5 performs a post pruning method, which is based on the estimation of the real error of the model, according to its apparent error, aiming to generalize the final model and avoid overfiting.

### 3.2 FuzzyDT

FuzzyDT is an algorithm to induce fuzzy decision trees based on the classic C4.5.The first steps are the definition of fuzzy partitions in the continuous attributes domains and the fuzzification of the attribute values. After that, the tree induction process is applied to generate the fuzzy rules. Algorithm 3.2.1 presents the main steps of FuzzyDT [2].

**Algorithm 3.2.1.** The FUZZYDT algorithm [2]

1. Define the fuzzy data base, i.e., the fuzzy granulation for the domains of the continuous features;
2. Replace the continuous attributes of the training set using the linguistic labels of the fuzzy sets with highest compatibility with the input values;
3. Calculate the entropy and information gain of each feature to split the training set and define the test nodes of the tree until all features are used or all training examples are classified with the same class label;
4. Apply a pruning process.

### 3.3 Fuzzy Unordered Rule Induction Algorithm: FURIA

FURIA, the Fuzzy Unordered Rule Induction Algorithm, was proposed in [10] as a modification and extension of the famous RIPPER algorithm.

The algorithm considers that a rule covers an example $x = (x_1, x_2, ..., x_n)$ if, and only if, the value of the attribute $x_i$ satisfy all predicates of the rules antecedent. Then, it orders the training examples according to the relative frequency of classes, from the least to the most frequent class. So, it learns rules for all classes, except for the last, which is the most frequent one. Once a rule is created, the examples covered by it are removed from the set of training examples.

The algorithm proceeds with the next class until there are no more examples in the training set or the last created rule is too much complex, according to a predefined measure.

Finally, RIPPER builds a default rule to the last class, which is the most frequent one. Intuitively, creating a default rule could be questionable, since it can privilege the most frequent class. One of the changes to this algorithm, that originated FURIA, is concerned with this default rule.

The main difference between FURIA and RIPPER is that FURIA infers fuzzy rules instead of crisp rules. Moreover, it does not order the training examples to infer the rules. Consequently, FURIA does not build a default rule, using a one-vs-rest decomposition to infer unordered rules.

When using an unordered rule set without default rule to classify a new instance, two problems can occur: First, a conflict may occur when the instance is equally well covered by rules from different classes. Second, it may happen that the instance is not covered by any rule. The first problem is rather unlikely to occur and, in case it still does, it is resolved by calculating the support of the rules and classifying the new instance as the class that occurs in the consequent of the rule which has higher support value. The second one is not so simple to resolve. For this, in [10], Cohen proposes a rule stretching method. The idea is to modify the rules in a local way so as to make them applicable to the instance that is been classified. It is done by replacing the rules by their minimum generalizations for the given instance. As proposed by [10], a generalization or stretching of a rule is obtained by deleting one or more of its antecedents, and it is minimal if it does not delete more antecedents than necessary to cover the instance. Thus, the minimal generalization of a rule is simply obtained by deleting all antecedents that are not satisfied by the instance.

Once all minimal generalizations are derived, FURIA re-evaluates each rule by its Laplace accuracy on the training data and then classify the instance by the rule with the highest evaluation.

## 4 Experiments

In this section we present the experiments developed, aiming to determine which of the three methods cited above (FuzzyDT, C4.5 or FURIA) gives better results for the soil classification problem.

The tests were performed using a real data set which instances were extracted from Brazilian System of Soil Information [11] and from researches on soil profiling, assigned by the Brazilian Institute of Geography and Statistics (IBGE) from Santa Catarina and by the Center of Agroveterinary Science of State University of Santa Catarina (CAV-UDESC).

To obtain the data set, a filter was applied to extract observations which follow the characteristics below:

- Altitude: upper than or equals to 600 meters;
- Soil's classes: Brown Latosol, Red Latosol , Red-yellow Latosol, Yellow Latosol, Brown Nitosol, Haplic Nitosol, Humic Cambisol, Haplic Cambisol;

- Location: from Paraná, Santa Catarina and Rio Grande do Sul states;
- Profile's sub-horizon B with largest amount of data.

While selecting the data, some attributes were ignored because they are not used neither to soils characterization nor classification.

Since there was only a small amount of observations from classes Red Latosol, Red-yellow Latosol, Yellow Latosol, Haplic Nitosol, Humic Cambisol and Haplic Cambisol, they were grouped into a single class which were named Other Latosols, Nitosols and Cambisols (OLNC).

After the preprocessing, the remaning data included instances from three possible classes:

(a) Brown Nitosol (BN)
(b) Brown Latosol (LB)
(c) Other Latosols, Nitosols e Cambisols (OLNC)

The characteristics of these soils were expressed by 25 attributes,described in Table 1 by means of the name, the type (discrete or continuous), the number of values in the case which the attribute is discrete and a brief descrition of each attribute.

The tests were carried out by using 10-fold Cross Validation. For C4.5 and FURIA algorithms, it was used the implementation of these algorithms available in the software WEKA [16] and for FuzzyDT, our own Java implementation. The parameters of the algorithms C4.5 and FURIA were maintained as the default ones and FuzzyDTs data fuzzyfication was done using partitions with three triangular fuzzy sets per attribute. The results, comprising the accuracy and number of rules generated by each method, are shown in Table 2. The rules format generated by each one of the algorithms are illustrated by the examples presented in Table 3.

As can be seen in Table 2, FuzzyDT obtains the best result in terms of accuracy, followed by FURIA and then C4.5. Concerning the number of rules, FuzzyDT generates the worse result, with a higher number than the other two methods. Although FURIA gives the lowest number of rules, the rules format do not favor comprehensibility of the system as a whole. While in the rules generated by C4.5 and Fuzzy DT it is possible to clearly identify both, the attribute which is been tested and its partition, with FURIA this recognition is not so simple to be done. This is mainly because the attribute values are represented by the parameters of trapezoidal membership function of its fuzzy sets. Besides that, analyzing the rule base constructed by FURIA, we realize that a different partition is generated for each attribute in each rule. This way, the fuzzy sets generated by the algorithm do not have a semantic meaning shared by all rules and the interpretability of the system is deteriorated.

**Table 1.** Attribute's characteristics

| **Attribute** | **Discrete/Continuous** | | **Description** |
|---|---|---|---|
| l_texture | Discrete | 3 | Level of texture of the soil. |
| l_structure | Discrete | 3 | Level of structure of the soil. |
| s_structure | Discrete | 3 | Structures size of the soil. |
| sh_structure | Discrete | 5 | Structures shape of the soil. |
| consistency_moist | Discrete | 3 | Level of consistency of the moist soil. |
| l_plasticity | Discrete | 3 | Level of plasticity of the soil. |
| l_tackiness | Discrete | 3 | Level of tackiness of the soil. |
| waxy | Discrete | 2 | Presence or absence of waxy and shiny appearance. |
| l_waxy | Discrete | 4 | Waxys level of the soil. |
| q_waxy | Discrete | 4 | Quantity of waxy of the soil. |
| l_distinctness | Discrete | 4 | Level of distinctness of the soil. |
| horizon_A | Discrete | 4 | Type of horizon A. |
| source_material | Discrete | 7 | Source material of the soil. |
| clay | Continuous | | Clay content of the soil. |
| cxc_clay | Continuous | | Cation exchange capacity of the clay. |
| fine_sand | Continuous | | Fine sand content of the soil. |
| grit | Continuous | | Grit content of the soil. |
| total_sand | Continuous | | Total sand content of the soil. |
| sulfuric_attack_$SiO_2$ | Continuous | | Si by sulfuric acid attack expressed by $SiO_2$. |
| sulfuric_attack_$Al_2O_3$ | Continuous | | Al by sulfuric acid attack expressed by $Al_2O_3$. |
| carbon_nitrogen | Continuous | | Carbon/Nitrogen. |
| $Fe_2O_3$_clay | Continuous | | $Fe_2O_3$/Clay content. |
| $Al_2O_3$_clay | Continuous | | $Al_2O_3$/Clay content. |
| $SiO_2$_clay | Continuous | | $SiO_2$/Clay content. |
| Ki_clay | Continuous | | Ki/Clay content. |

**Table 2.** Tests' results to C4.5, FURIA and FuzzyDT algorithms

| **C4.5** | | **FURIA** | | **FuzzyDT** | |
|---|---|---|---|---|---|
| Accuracy | # of Rules | Accuracy | # of Rules | Accuracy | # of Rules |
| 82.85 | 50 | 83.99 | 20 | 92.99 | 104 |

**Table 3.** Example of rules format generated by the three methods

| **Method** | **Example of Rule** |
|---|---|
| C4.5 | grit $\leq 120$ **and** $SiO_2$_clay $\leq 0.088222$: BN |
| FURIA | (source_material = Sao_Bento) **and** (sulfuric_attack_$Al_2O_3$ **in** $[-\infty, -\infty, 227, 232]$) $\geq$ classe=OLNC |
| FuzzyDT | **IF** source_material **IS** 4 **AND** grit **IS** low **THEN** CLASS **IS** 3 |

## 5 Conclusion

The best use of the different types of soil depends on a proper classification. This is not an easy task since it implies very subjective expert opinions. Aiming at solving this problem, we proposed and tested the use of a fuzzy decision tree approach, named FuzzyDT to build a fuzzy classification system which deals with the problem of soil classification. We compared the generated FuzzyDT with two other classification systems, obtained from the algorithms C4.5 and FURIA. Analysing the results, it is possible to observe that FuzzyDT reaches the highest accuracy but generates the highest number of rules. In spite of that,, its rules are interpretable, following the format of standard fuzzy rules. FURIA and C4.5 obtained very similar results with respect to accuracy while FURIA generates the lower number of rules. Nevertheless, it generates rules that are not interpretable, once for each rule, a different partition for each attribute is generated, which implies that the fuzzy sets generated by the algorithm are not interpretable.

In the future work we intend to investigate techniques to be applied on the set of fuzzy rules generated by FuzzyDT, to reduce the number of rules, while still preserving the good accuracy obtained.

## References

1. Santos, H., et al.: Brazilian System of Soil Classification. Embrapa Solos, Rio de Janeiro, RJ, Brazil (2013) (in Portuguese)
2. Cintra, M.E., Monard, M.C., Camargo, H.A.: A Fuzzy Decision Tree Algorithm Based on C4.5. Mathware & Soft Computing 20, 56–62 (2013)
3. Chang, R.L.P., Pavlidis, T.: Fuzzy decision tree algorithms. IEEE Transactions on Systems Man and Cybernetics 7, 28–35 (1977)
4. Janikow, C.Z.: Fid4.1: an overview. In: Proceedings of NAFIPS 2004, pp. 877–881 (2004)
5. Kazunor, U., Motohide, S.: Fuzzy C4.5 for generating fuzzy decision trees and its improvement. Faji Shisutemu Shinpojiumu Koen Ronbunshu 15, 515–518 (2001) (in Japanese)
6. Olaru, C., Wehenkel, L.: A complete fuzzy decision tree technique. Fuzzy Sets and Systems 138(2), 221–254 (2003)
7. Tokumaru, M., Muranaka, N.: Kansei impression analysis using fuzzy c4.5 decision tree. In: International Conference on Kansei Engineering and Emotion Research (2010)
8. Cintra, M.E., Meira, C.A.A., Monard, M.C., Camargo, H.A., Rodrigues, L.H.A.: The use of fuzzy decision trees for coffee rust warning in Brazilian crop. In: 11th International Conference on Intelligent Systems Design and Applications, vol. 1, pp. 1347–1352 (2011)
9. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco (1993)
10. Huhn, J., Hullermeier, E.: Furia: An algorithm for unordered fuzzy rule induction. Data Mining and Knowledge Discovery (2009)

11. Oliveira, J.B.: Soil classification and its agricultural and nonagricultural use (2013) (in Portuguese), http://jararaca.ufsm.br/websites/dalmolin/download/textospl/classif.pdf (accessed in December 30, 2013)
12. Almeida, J.: Brief History of the Subtropical Bruno Soils in Brazil. VIII RCC Brazilian Meeting of Soil Correlation and Classification. Embrapa Solos (2008) (in Portuguese)
13. Cohen, W.W.: Fast effective rule induction. In: Proceedings of the Twelfth International Conference on Machine Learning, pp. 115–123. Morgan Kaufmann (1995)
14. Furnkranz, J., Widmer, G.: Incremental Reduced Error Pruning. In: Proceedings of the 11th International Conference on Machine Learning (ML 1994), pp. 70–77. Morgan Kaufmann, New Brunswick (1994)
15. Furnkranz, J.: Separate-and-conquer rule learning. Artificial Intelligence Review 13(1), 3–54 (1999)
16. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. SIGKDD Explorations 11(1) (2009)
17. Cordón, O., del Jesús, M.J., Herrera, F.: A proposal on reasoning methods in fuzzy rule-based classification systems. Int. J. Approx. Reasoning 20(1), 21–45 (1999)
18. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
19. Shannon, C.: A mathematical theory of communication. Bell System Technical Journal 27 (1948)
20. Martinez, C.V.O., Souza, V.F.: Importance of the soils classification in the Brazilian system and the ability of land use of rural properties for their sustainable management. In: International Meeting of Cientific Production Cesumar (2009) (in Portuguese)

# Towards the Use of Sequential Patterns for Detection and Characterization of Natural and Agricultural Areas

Fabio Guttler[1], Dino Ienco[1], Maguelonne Teisseire[1], Jordi Nin[2], and Pascal Poncelet[3]

[1] IRSTEA, UMR TETIS, Montpellier, France
{fabio.guttler,dino.ienco,maguelonne.teisseire}@teledetection.fr
[2] Barcelona Supercomputing Center (BSC)
Universitat Politècnica de Catalunya (BarcelonaTech), Barcelona, Catalonia, Spain
nin@ac.upc.edu
[3] LIRMM, CNRS, Montpellier, France
pascal.poncelet@lirmm.fr

**Abstract.** Nowadays, a huge amount of high resolution satellite images are freely available. Such images allow researchers in environmental sciences to study the different natural habitats and farming practices in a remote way. However, satellite images content strongly depends on the season of the acquisition. Due to the periodicity of natural and agricultural dynamics throughout seasons, sequential patterns arise as a new opportunity to model the behaviour of these environments. In this paper, we describe some preliminary results obtained with a new framework for studying spatiotemporal evolutions over natural and agricultural areas using $k$-partite graphs and sequential patterns extracted from segmented Landsat images.

**Keywords:** Temporal Patterns, Data Mining and Remote Sensing.

## 1 Introduction

Several regions over the earth are composed of complex landscapes with regard to land cover and land use. Outside urban areas, we commonly observe mosaics of natural, semi-natural and agricultural areas. In Europe, mapping and monitoring those areas became a major issue and several procedures have been established for guiding and controlling such tasks. With regard to natural areas we can cite the Habitats Directive (92/43/EEC) and the associated Natura 2000 network of protected sites. In this context, member states must report the conservation status of the habitats within their territory every six years [5]. Concerning agriculture, Land Parcel Identification Systems (LPIS) emerged since 1992 (Council Reg. No 3508/1992). LPIS is used as a reference for annual declaration by farmers, for administrative and cross-checks, and on the spot controls (including control with remote sensing) [13].

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 97–106, 2014.

The detection of natural and agricultural areas from remote sensing data is a key point for land cover mapping. Image classification is a widespread method for mapping land cover. The overall objective is to categorize all pixels in an image into land cover classes [9]. Usually, image classification is based on the differential spectral response of land surfaces (radiance values recorded at pixel level). However, most of classifications are based on single image and present lot of bias in their results, or require extensive ground truth data in order to attempt a high accuracy.

Temporal pattern recognition can furnish complementary information to feature identification. Actually, natural vegetation and agricultural crop present distinct evolutions during phenological cycles and growing seasons. In that light, multi-date imagery can enhance the pertinence and the accuracy of land cover detection. Time series analysis comprises methods for analysing temporal data [12], several images in our case, in order to extract meaningful statistics and other characteristics of the data. Temporal data allows researchers to create a data model for analysing past values and forecasting future ones [11]. Specifically, this work is focus on natural and agricultural areas modelling over time. Sequential pattern mining is the part of time series analysis concerned with finding statistically relevant patterns between data examples where the values are delivered in a different time moments [10].

Optical remote sensing, such as Landsat images are commonly used in environmental researches. Image processing techniques are usually grouped in Pixel based analysis or Object based image analysis (OBIA). Several time series pixel-based approaches have been proposed, but OBIA studies rarely uses multi-temporal data [12].

In this paper we combine OBIA with sequential pattern mining to create a $k$-partite graph to represent natural and agricultural areas. Our approach starts segmenting multi-date satellite images over a same area. From these segmentations, a $k$-partite graph is built considering objects as nodes and representing object image overlap as edges. Later, in order to determine the optimum bounding box of each object and study its dynamics, we propose the creation of $k$-partite subgraphs. Finally, subgraphs are supplied to an expert in order to be categorized.

### 1.1 Paper Organisation

This paper is organised as follows: Firstly, we introduce some concepts about optical remote sensing in Section 2. Section 3 describes $k$-partite graphs and sequential pattern representation. Our technique to maximise image coverage is introduced in Section 4. Experimental findings about agricultural and natural areas are depicted in Section 5. Section 6 concludes and draws future works.

## 2 Remote Sensing Satellite Images

In general we talk about remote sensing when the acquisition of data is done without making any physical contact. This is the case of Earth observation plat-

forms such as Landsat. The optical sensors onboard Landsat acquires multispectral images (composed by multiple bands), each band representing a portion of the electro-magnetic spectrum. For this study, we used Landsat Thematic Mapper (TM) images without the thermal band. Remote sensing images are usually characterized by different resolutions and other technical characteristics, in our case they are as follows:

- **Spatial Resolution.** The size of a pixel in a raster image is 30x30 meters.
- **Spectral Resolution.** The spectral bands are six: blue, green, red, near infrared (NIR) and two parts of short wavelength infrared (SWIR-1 and 2).
- **Radiometric Resolution.** The sensor is able to distinguish is 256 intensities of radiation (8 bits).
- **Swath Width.** The scene size is a ground square of about 185 x 185 km.

Apart from these parameters another important issue in time-series studies is the frequency of revisits by the satellite. The revisit time of Landsat is of 16 days, but in practice the number of useful images is lower, mostly due to unsuitable whether conditions during satellite acquisition. For this study, we selected six images covering of the same geographic area (the Natura 2000 Lower Aude Valley site, located in south of France) between February and September 2009.

To avoid errors and misplacement over time, all images were already georeferenced, as well as radiometrically and atmospherically corrected by CESBIO-CNES [6]. Additional fine spatial positioning corrections were applied in order to keep the spatial shift between all the time stamps less than one pixel.

### 2.1 Segmentation

Image segmentation is a fundamental step in OBIA and its consists in merging pixels into objects clusters [3]. Objects (or segments) are regions generated by one or more criteria of homogeneity in one or more dimensions of a feature space [4]. The aim of segmentation is to create a new representation of the image more meaningful and easier to analyse. This approach is similar to human visual interpretation of digital images, which works at multiple scales and uses colour, shape, size, texture, pattern and context information [9].

Image segmentation results in a set of objects that collectively cover the entire image without any overlapping. With respect to the homogeneity criteria, adjacent objects are expected to be significantly different between them. In this work we use the software eCognition Developer 8.8.1 for image segmentation (multiresolution segmentation algorithm). Only the pixels within the boundaries of the Lower Aude Valley Natura 2000 site (4,842 ha) were used for segmentation and further processing steps. Nine raster layers (radiometric or 'colour' information) were used simultaneously for image segmentation. Six of them correspond to the Landsat spectral bands and the other are spectral indices. Spectral indices are commonly used in remote sensing as they can be helpful for detecting and characterizing some specific features, like vegetation, soil, water, etc. We used the Normalized Difference Vegetation Index (NDVI) [8], the Normalized Difference Water Index (NDWI) [7] and the Visible and Shortwave Infrared Drought

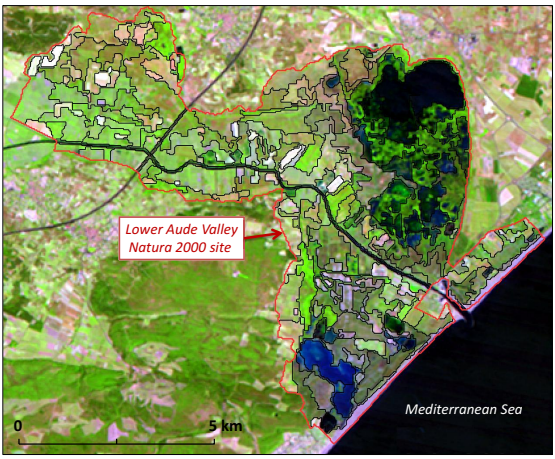



**Fig. 1.** Segmentation example representing the time stamp of 10 July 2009

Index (VSDI) [14]. In order to obtain objects of interest related to natural and agricultural areas, we conceived a segmentation rule-set composed of 3 main steps:

1. Medium-coarse segmentation to delineate general zones (colour and shape components combined but "color > shape"): about 170-200 objects
2. Very fine segmentation focused on colour component: about 6,000 objects
3. Medium-fine segmentation with balanced weights for colour and shape components: about 500-600 objects

This process was independently applied for each Landsat image. The last segmentation level (3) was then exported from each time-stamp and used as input for the subsequent processing steps. Figure 1 illustrates the segmentation layer obtained for the time stamp of 10 July 2009.

## 3 Sequential Patterns and *k*-partite Graphs

This section describes the traditional sequential pattern mining problem and high-lights the need for a specific way to handle remote sensing temporal information by using $k$-partite graphs.

### 3.1 *k*-partite Graphs

A $k$-partite graph is a graph $G = (V, E)$ with vertex set $V$ and edge set $E$, whose graph vertices $V$ can be partitioned into $k$ disjoint sets $V_k$ so that no two vertices within the same set are adjacent. In this work, we assume that any edge $e_{ij}$ has a weigh equal to $w_{ij}$ where $i$ and $j$ correspond to two vertexes $v_i \in V_i$ and $v_j \in V_j$ in two consecutive sets (layers).

### 3.2 Sequential Patterns

Sequential patterns were introduced in [2] and in general, they are considered as an extension of the concept of frequent itemset [1] having timestamps associated to items. Sequential pattern mining aims at extracting sets of items commonly associated over time.

The problem of generating and mining sequential patterns as a $n$-bipartite graph is defined as follows.

Let $X$ be a set of distinct items. An itemset is a subset of items, denoted by $I = (i_1, i_2, \ldots, i_n)$, for $1 \leq j \leq n, i_j \in X$. A $k$-partite graph is built as an ordered list of itemsets, denoted by $\langle I_1, I_2, \ldots, I_k \rangle$, where $I_i \in X$ for $1 \leq i \leq n$ and correspond with each $V_i$ nodes set.

Once graph nodes are defined, it is required to define the edges and their corresponding weights. To do that, a function $f(I_i, I_j) = w_{ij}$ is defined, if $f(I_i, I_j) < 0$ it means that there is one edge between a pair of nodes $v_i$ and $V_j$ with weight equal to $w_{ij}$.

### 3.3 Representation of Natural and Agricultural Areas

In order to represent natural and agricultural area evolutions over time, it is proposed to create a sequential pattern by using a $k$-partite graph representation in the following way:

1. Segmented images are the different itemsets $\langle I_1, I_2, \ldots, I_k \rangle$, note that, the image timestamp provide us the temporal information for creating the different nodes sets and therefore the $k$-partite graph.
2. Objects extracted from the segmented images allow us to generate the items $i_l$ of each itemset $I_j$.
3. To create the edges, the function $f(I_i, I_j) = |p_{i_l} \cap p_{i_{l+1}}|/|p_{i_l} \cup p_{i_{l+1}}|$ is calculated, where $p_{i_l}$ stands for the list of pixels of the item $i_l$.

Within this global $k$-partite graph, it is possible to extract a concrete area of interest evolution creating a sub-graph chosen one or more items. All this process is illustrated in Figure 2.

## 4 Coverage Algorithm

In this paper we have implemented the following algorithm to (completely) cover one region of interest with sequential patterns: firstly, we detect the previously unprocessed areas of the image where frequent changes in the pixel values are produced. After that, we select from the first image the objects included in the previously selected image area. Later, we generate a sequential pattern and mark this part of the image as processed and recompute object variability. The algorithm stops when the region of interest has been completely processed.

Specifically, we implemented the above coverage algorithm as it is described in Algorithm 1. Firstly, we create a bounding box list containing all the objects of all the images (lines 2-5), at the same time, we compute the variability of

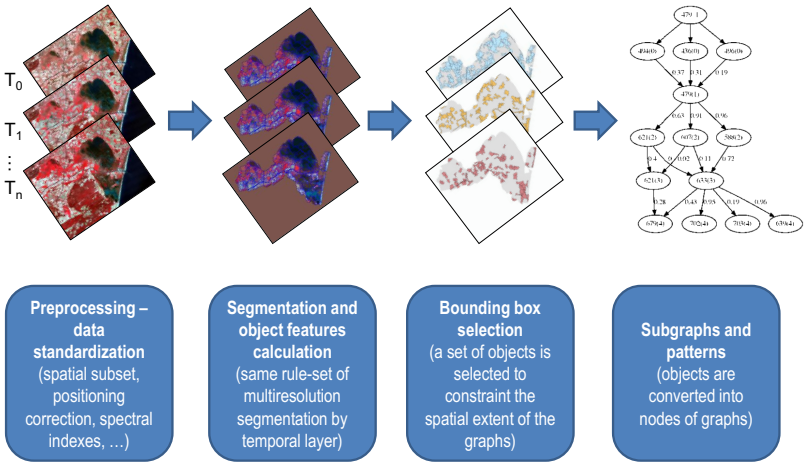



**Fig. 2.** Overall Sequential Patterns Extraction Process

**Algorithm 1.** BoundingBox Selection

```
   Data: g: main graph
   Result: p: patterns list
1  begin
2      bbList = new BBList;
3      foreach layer l ∈ g do
4          foreach node n ∈ l do
5              bb = new bb(n, l, importance(n));
6              bbList = bbList ∪ bb;
7      while !empty(bbList) do
           /* 1. Find the corresponding objects in Layer 0 for the most important bounding box */
8          box = bbList.MostImportant();
9          n0 = g.getNodesByBB(0, box);
           /* 2. Create the pattern for the selected object with the selected bounding box computing the overlap always
              with the selected nodes in Layer 0 */
10         p = p ∪ newPattern(n0, box);
           /* 3. Remove all the objects contained in the patterns from the importance of the BB */
11         bbList.updateImportance(p);
```

each bounding box (object). Bounding box importance is the amount of other objects overlapping with at least one pixel with the current bounding box, this computation is described in Algorithm 2.

Once the bounding box list is created, all objects are ranked with regards their importance. The importance of a bounding box corresponds with the number of objects in any layer overlapping with at least one pixel with the bounding box. After that, the most important bounding box is selected (line 7) and recover the objects overlapping with such bounding box in the first image (line 8). Then, a new pattern is generated as it is described in Section 3.2 (line 9). After that, we update bounding boxes importance, subtracting the processed objects from the importance of each bounding box, if bounding box importance is equal to 0, it is removed from the list (Algorithm 3). This loop is repeated until the bounding box list is empty.

**Algorithm 2.** Importance

```
Data: g: main graph, b:bounding box
Result: i: importance
1 begin
2     i = 0;
3     foreach layer l ∈ g do
4         foreach node n ∈ l do
              /* if one pixel of node n overlaps
                 with g                          */
5             if n.overlap(b) then
6                 i++;
```

**Algorithm 3.** Updateimportance

```
Data: p: pattern
Result: bb:bounding box list
1 begin
2     foreach boundingBox b ∈ bb do
3         foreach layer l ∈ p do
4             foreach node n ∈ l do
                  /* if one pixel of node n
                     overlaps with b          */
5                 if n.overlap(b) then
6                     b.i − −;
7                 if b.i == 0) then
8                     bb.remove(b);
```

Figure 3 shows the obtained coverage within the perimeter of our study area. Each polygon represents the spatial extent of a specific bounding box, what means that the associated pattern has a bigger spatial coverage taking into account all temporal layers. Polygon color indicates from which time stamp the bounding box was extracted (as assigned on the legend). In total 331 bounding box were selected (about 50 - 60 per timestamp), it corresponds to an overall coverage of 93.56% of the study area.

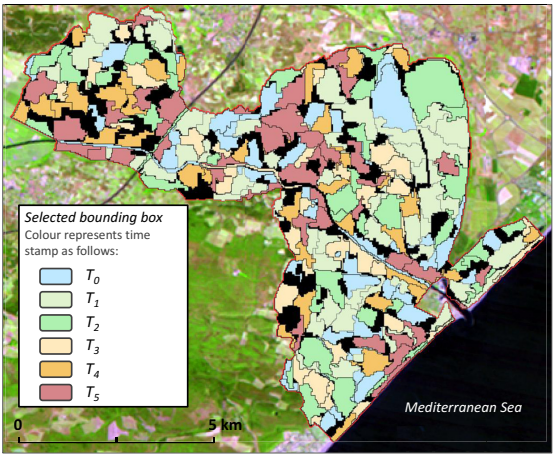



**Fig. 3.** Obtained spatial coverage

## 5 Experiments

In order to evaluate the pertinence of our approach, graphical results were analyzed by a remote sensing expert with field knowledge of the study area. For this analysis, all subgraphs (331 in total) were plotted with an associated unique ID allowing to locate each bounding box in an Geographic Information System (GIS). The expert preselected (without taking into account subgraph results)

about 20 regions of interest (ROI) containing both natural and agricultural areas. Then, the subgraphs corresponding to this set of ROI were identified and analyzed. The following subsection details the main differences between the graphs representing spatio-temporal evolutions of agricultural and natural areas within our study area.

### 5.1 Pattern Analysis Examples

**Agricultural Area.** In general, agricultural areas present less complex subgraphs as the fields are usually rectangular and the boundaries of the objects remain very similar from one time stamp to the next one. Even if subgraph structure is quite simple (i.e. Figure 4 - subgraph 1), radiometric information varies sharply throughout the time series. Subgraph 1 (Figure 4) illustrates the temporal pattern of a winter culture (cereal crop in this example). The parcel presents an important augmentation of the NDVI from late winter up to spring, which corresponds to the growing season. The crop is harvest in late spring or early summer, generating a brutal change in the spectral response (between T2 and T3). Afterwards, the field is characterized by bare soil, which remains globally stable until the end of the time series.

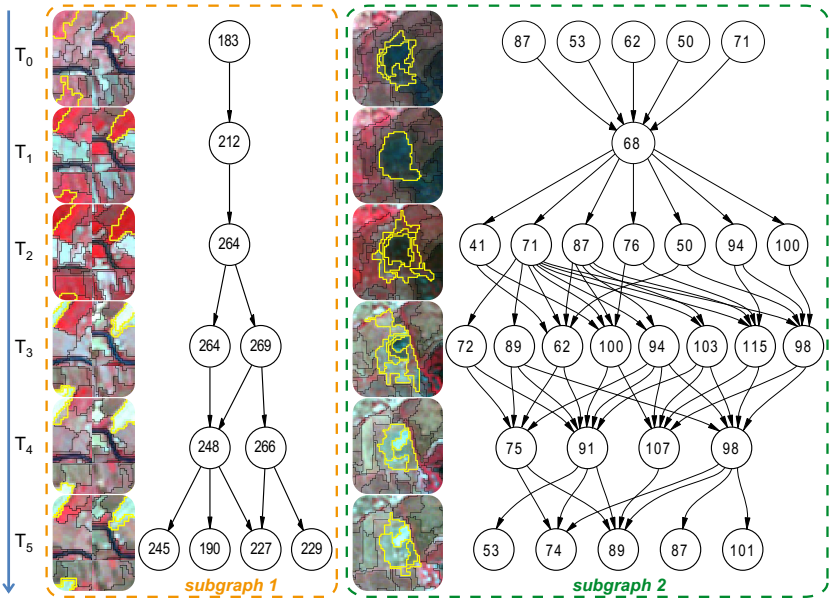



**Fig. 4.** Temporal Pattern examples: Agricultural area (subgraph 1) and Natural area (subgraph 2)

**Natural Area.** In natural areas the boundaries are usually much smoother than in agricultural fields and the objects rarely have rectangular shapes. As a consequence, graph structure tends to be more complex and can englobe several objects per layer. Such kind of graph can also show some drastic radiometric changes, but in general it concerns only a small portion of the area covered by the bounding box. In other words, natural habitats are characterized by more gradual and smoothed evolutions. Subgraph 2 of figure 4 illustrates the temporal pattern of a temporary lake. Usually flooded during winter and spring, this lake becomes progressively dry starting from early summer. Aquatic natural habitat is gradually replaced by some pioneer terrestrial communities, dominated by annual *Salicornia sp.* or *Suaeda sp.* Depending mainly on the local variations of soil salinity, the vegetation will cover the former flooded area in different ways. First developments of such particular habitats can be detected since T2 and they should subsist up to the next flood event (usually during late autumn or winter).

### 5.2 Discussion

Throughout one year, spatio-temporal evolutions of agricultural and natural areas are enough dissimilar and it is possible to an expert to separate them by analysing the time series of satellite images. In a more automatic way, the subgraphs generated by our approach synthesizes complex spatio-temporal evolutions and can be useful for this kind of time-consuming task. Actually, subgraph structure associated to temporal behaviour of object's radiometry provides complementary and pertinent information allowing detailed analysis.

At this stage, our coverage algorithm is not able to cover 100% of the study area throughout the time series. Most of times, small objects are not included in the subgraphs. However, this issue should be improved as the spatial gaps concerns sometimes also medium-size objects representing interesting natural or agricultural areas. Another point to improve is spatial redundancy. We verified that some subgraphs presents high rates of similar objects, what means that the same area is covered by more than one subgraph. Improving the bounding box selection should reduce such spatial redundancy.

## 6 Conclusions

In this paper we have described a complete framework for studying evolving natural and agricultural areas using satellite images information and $k$-partite graph sequential patterns. We have shown that using our approach is possible to cover a great part of the study area (93.56%) and to analyze in detail a concrete region of interest. We have also verified with one domain expert that the obtained sequential patterns are meaningful. As a future work, we would like to develop some clustering algorithms for $k$-partite graphs to help the expert to post-process the results providing some groups of patterns with similar behaviour, instead of individuals patterns without relation among them.

**Acknowledgments.** This work was supported by public funds received in the framework of GEOSUD, a project (ANR-10-EQPX-20) of the program "Investissements d'Avenir" managed by the French National Research Agency. This work was also partially supported by the Ministry of Science and Technology of Spain under contract TIN2012-34557 and by the BSC-CNS Severo Ochoa program (SEV-2011-00067).

## References

1. Agrawal, R., Imielinski, R., Swami, A.: Mining association rules between sets of items in large databases. SIGMOD Records 22(2), 207–216 (1993)
2. Agrawal, R., Srikant, R.: Mining sequential patterns. In: ICDE (1995)
3. Baatz, M., Hoffmann, C., Willhauck, G.: Progressing from object-based to object-oriented image analysis, ch. 2. Lecture Notes in Geoinformation and Cartography, pp. 29–42. Springer, Heidelberg (2008)
4. Blaschke, T.: Object based image analysis for remote sensing. ISPRS J. Photogramm. 65(1), 2–16 (2010)
5. Vanden Borre, J., Paelinckx, D., Mücher, C.A., Kooistra, L., Haest, B., De Blust, G., Schmidt, A.M.: Integrating remote sensing in natura 2000 habitat monitoring: Prospects on the way forward. J. Nat. Conserv. 19(2), 116–125 (2011)
6. Hagolle, O., Huc, M.: Séries temporelles de produits landsat de niveau 2a: Manuel de l'utilisateur des données produites au cesbio (2011)
7. Jackson, T.J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., Doriaswamy, P., Hunt, E.: Vegetation water content mapping using landsat data derived normalized difference water index for corn and soybeans. Remote Sens. Environ. 92(4), 475–482 (2004)
8. Rouse Jr, J., Haas, R., Schell, J., Deering, D.: Monitoring vegetation systems in the great plains with erts. NASA Special Publication 351, 309 (1974)
9. Lillesand, T.M.: Remote Sens. and Image Interpret. John Wiley & Sons (2006)
10. Nin, J., Laurent, A., Poncelet, P.: Speed up gradual rule mining from stream data! a b-tree and owa-based approach. JIIS 35(3), 447–463 (2010)
11. Nin, J., Torra, V.: Towards the evaluation of time series protection methods. Inform. Sciences 29(11), 1663–1677 (2009)
12. Petitjean, F., Kurtz, C., Passat, N., Gançarski, P.: Spatio-temporal reasoning for the classification of satellite image time series. Pattern Recogn. Lett. 33(13), 1805–1815 (2012)
13. P. Trojáček and R. Kadlubiec. Detailed mapping of agricultural plots using satellite images and aerial orthophoto maps (2004)
14. Zhang, N., Hong, Y., Qin, Q., Liu, L.: Vsdi: a visible and shortwave infrared drought index for monitoring soil and vegetation moisture based on optical remote sensing. Int. J. Remote Sens. 34(13), 4585–4609 (2013)

# Application of $E^2M$ Decision Trees to Rubber Quality Prediction

Nicolas Sutton-Charani[1,2], Sébastien Destercke[1], and Thierry Denœux[1]

[1] Université Technologie de Compiègne, UMR 7253 Heudiasyc
60203 COMPIEGNE Cedex FRANCE
{name.surname}@hds.utc.fr
http://www.hds.utc.fr/
[2] CIRAD IATE, 2, place Pierre Viala F-34060 Montpellier Cedex 02 FRANCE
F-34392 Montpellier, France

**Abstract.** In many applications, data are often imperfect, incomplete or more generally uncertain. This imperfection has to be integrated into the learning process as an information in itself. The *$E^2M$ decision trees* is a methodology that provides predictions from uncertain data modelled by belief functions. In this paper, the problem of rubber quality prediction is presented with a belief function modelling of some data uncertainties. Some resulting *$E^2M$ decision trees* are presented in order to improve the interpretation of the tree compared to standard decision trees.

**Keywords:** classification, decision trees, rubber quality, hevea, belief functions, algorithm EM.

## 1 Introduction

Learning a classifier from uncertain data necessitates an adequate modelling of this uncertainty, however learning with uncertain data is rarely straightforward. As data uncertainty is of epistemic nature, the standard probabilistic framework is not necessarily the best framework to deal with it. More general frameworks have therefore been proposed [1–3] that provide more adequate model for this type of uncertainty. Different classifier learning techniques [4–6] using these models have then been developed.

In this paper, our goal is to learn a model from agronomic data. More precisely, we want to predict natural rubber quality from data concerning latex culture and natural rubber maturation. Generally speaking, uncertain measurements and expert assessments are common in agronomy and life science, mainly due to field and economy constraints. They are therefore domains where data uncertainty happens a lot. We retain the belief-function theory [2, 7], as it is flexible enough to model a large variety of data uncertainties. The chosen classifier is the $E^2M$ decision tree [8], for it is usually efficient and interpretable (an essential feature for agronomic experts).

After a short overview on the necessary background in Section 2, we detail in Section 3 the application context as well as the uncertainty models we used. We conclude Section 3 by comparing the results of the obtained $E^2M$ decision trees with classical ones.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 107–116, 2014.

## 2 Background

We briefly recall the elements needed in the application.

### 2.1 Formalism

As in any classification problem, the aim is to predict a class label $Y$ from a set of attributes (or features) $X$. The classifier is learnt on a learning dataset $LD$ containing samples of $(X,Y)$. The classifier is then evaluated by measuring its accuracy on a test dataset $TD$, comparing the predicted class labels with the real ones.

The attributes $X = (X^1,\dots,X^J)$ take their values on $\Omega_X = \Omega_{X^1} \times \cdots \times \Omega_{X^J}$, the class $Y$ on $\Omega_Y = \{\omega_1,\dots,\omega_K\}$. That is, $K$ different classes are predicted using $J$ different attributes (either categorical or real-valued).

A *precise* dataset containing $N$ samples is a set of observations of $(X,Y)$ and is denoted by

$$D = \begin{pmatrix} x_1, y_1 \\ \vdots \\ x_N, y_N \end{pmatrix} = \begin{pmatrix} x_1^1,\dots,x_1^J, y_1 \\ \vdots \\ x_N^1,\dots,x_N^J, y_N \end{pmatrix}.$$

Samples are here assumed to be i.i.d (independant and identically distributed).

### 2.2 Belief-Function Theory

The theory of belief functions ($TBF$), also called evidence theory or Dempster-Shafer theory was first presented by Dempster [2] in a statistical approach. The very basis of the $TBF$ is here presented, with a special focus on the evidential likelihood proposed by Denoeux [9].

**Generalities.** Assume we have an uncertain observation of a variable $W$ defined on a finite space $\Omega_W$. We model this observation by a *belief mass* $m^W : 2^{\Omega_W} \to [0,1]$ verifying $\sum_{B\in 2^{\Omega_W}} m^W(B) = 1$. We assume here that $m^W(\emptyset) = 0$. A *focal element* $A \in 2^{\Omega_W}$ is a set such that $m^W(A) > 0$. From this mass, the belief and plausibility functions are defined by:

$$Bel^W(A) = \sum_{B\subseteq A} m^W(B), \qquad Pl^W(A) = \sum_{B\cap A\neq\emptyset} m^W(B)$$

$Bel^W(A)$ measures the amount of information that implies $W \in A$, and is a measure of certainty, while $Pl^W(A)$ measures the amount of information that does not conflict with $W \in A$, and is a measure of plausibility. We naturally have $Bel^W(A) \leq Pl^W(A)$ with the two being equal in the specific case of probabilities.

The particular cases of precise data, imprecise data, missing data, probabilities and possibilities can all be modelled by belief functions:

$$\begin{aligned}
\text{precise data}: &\quad m^W(\{w\}) = 1 \\
\text{imprecise data}: &\quad m^W(A) = 1 \\
\text{missing data}: &\quad m^W(\Omega_W) = 1 \quad (\textit{complete ignorance}) \\
\text{probabilities}: &\quad m^W(A) > 0 \quad \textit{if} \quad |A| = 1 \\
\text{consonnant mass functions}: &\quad m^W(A) > 0 \text{ and } m^W(B) > 0 \quad \textit{only if} \quad A \subset B \text{ or } B \subset A
\end{aligned}$$

In our classification context, an evidential dataset $ED$ will be of the form

$$ED = m^{X,Y} = \begin{pmatrix} m_1^{X,Y} \\ \vdots \\ m_N^{X,Y} \end{pmatrix} = \begin{pmatrix} m_1^{X^1} & \cdots & m_1^{X^J} & m_1^Y \\ \vdots & \ddots & \vdots & \vdots \\ m_N^{X^1} & \cdots & m_N^{X^J} & m_N^Y \end{pmatrix}$$

where $m_i^{X,Y}$ describes the $i^{th}$ sample with its uncertainty.

**Evidential Likelihood.** Assume now we want to fit a parametric model with parameter $\theta$ to the data. Likelihood maximisation often provides a good estimator $\hat{\theta}$ of the unknown parameter $\theta$. When data are uncertain, the likelihood can be re-written in the following way:

$$\begin{aligned}
\textit{precise likelihood:} \quad L(\theta; w) &= P_\theta(W = w) \\
\textit{imprecise likelihood:} \quad L(\theta; A) &= \sum_{w \in A} L(\theta; w) \\
\textit{evidential likelihood:} \quad L(\theta; m^W) &= \sum_{A \subseteq \Omega_W} m^W(A_i) L(\theta; A_i)
\end{aligned} \tag{1}$$

As shown in [9], this evidential likelihood can be maximised by the adaptation of the $EM$ algorithm to belief functions: the $E^2M$ algorithm. This algorithm is quite similar to the $EM$ and is guaranteed to converge towards a local maximum. The main difference is at the Expectation step ($E$), where the measure used to compute the expectation is the conjunctive combination of $P_\theta$ and $m^W$.

### 2.3 Decision Trees

Decision trees are basic classifiers widely used in many areas such as machine learning, statistics, data mining etc. They are usually built from precise datasets by partitioning the attribute space in a set of leaves, each leave being thus attached to some conditions on the attribute values and to a predicted class.

As each leave is characterized by the proportion of the population "falling" into it as well as the frequencies of the class within this population, a decision tree can be viewed as a multinomial mixture whose parameters are the leaves probabilities (corresponding to the mixture coefficients) and the class probabilities inside the leaves (multinomial).

*Example 1.* Let us consider a data set with $N = 12$ data items, $J = 2$ attributes and $K = 3$ classes. Spaces are described as follows:

$$\Omega_{X^1} = [1,10], \quad \Omega_{X^2} = [13,150], \quad \Omega_Y = \{a,b,c\}.$$

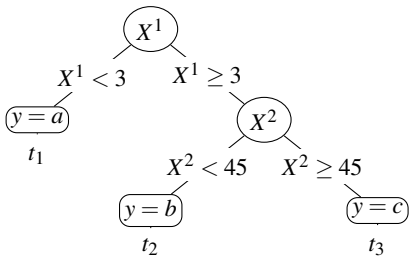



**Fig. 1.** Decision tree illustration

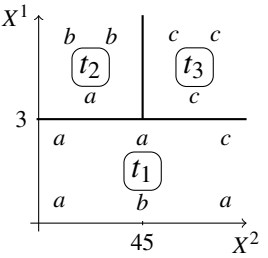



**Fig. 2.** Partitioned attribute space

Figures 1 and 2 respectively represent a possible decision tree and its corresponding partition of the attribute space with the learning dataset represented in it. To leaf $t_2$ are associated the values $A_2^1 \times A_2^2 = [3,10] \times [13,45[$, its estimated probability is $3/12$ and the class estimated probabilities are ($1/3$, $2/3$, 0), and the prediction for any test data item in $t_2$ (for example, $x^1 = 6$ and $x^2 = 30$) will be $y = b$.

The growing of a decision tree is recursive, and aims at separating classes inside leaves. This *separation* is usually measured by an impurity measure $i$ computed from the leaves probabilities and the class distributions inside leaves. From a root node containing all the samples of the learning dataset, all possible splits (on all attributes and for all values) are explored, and the one with the highest purity gain (i.e., highest impurity reduction) is chosen. Then, for each newly created leave and the sample "falling" into it, the process is iterated until some stopping criteria is reached.

### 2.4 The $E^2M$ Decision Trees

To learn decision trees from uncertain data (where potentially both attributes and classes can be uncertain), we proposed [8] to learn the multinomial mixture of the tree through the $E^2M$ algorithm. Périnel [10] proposed a similar idea, yet only dealt with uncertain attributes and probabilistic data. We refer to [8] for technical details about the learning process of $E^2M$ decision trees, as well as for some experiments on benchmark data comparing $E^2M$ and *CART* decision trees and showing the potential interest of $E^2M$ decision trees in terms of accuracy in noisy environments.

## 3 Application: The Rubber Quality Problem

We first present the application and two uncertainty models we used on the data, before showing some interesting results. Note that the uncertainty models could be re-used in similar situations.

### 3.1 Problem Description

Natural rubber is the result of a milky fluid transformation: the *latex* that is extracted from the *hevea* tree. Compared to synthetic rubber, natural rubber presents some unique physical properties but suffers from a pretty high variability in terms of quality, that experts are not fully able to explain and to control.

In order to better control this quality, one of the biggest rubber company has put some effort to study this variability in one of its plantations by collecting cultural data. This plantation is located in a part of Brazil where the climate has the particularity to be more *variable* than in the rest of Brazil (the natural origin of the *hevea* tree is in the Brazilian forests). However, due to the size of the plantation (approximatively 70 hectars) and various factors (e.g., untracked delay of collection/process due to weather conditions, tanks mixing productions of many parcels), some variables are subject to high uncertainties.

Data are constituted of many variables summarised in Table 1 (no Unit means a dimensionless variable). Meteorological data may influence the latex during three different periods: the latex fabrication by the tree (one week before tapping), the tapping day during which latex is collected, and the latex maturation in tanks (five days). For the temperature and the relative humidity, the minimum, median and maximum values are computed for each day.

The data set contains 3053 examples described by 106 attributes. The quality is measured by the $P_{30}$ index which is an elasticity index. In order to use the $E^2M$ decision trees methodology, the $P_{30}$ was discretised into 5 equiprobable classes. This discretisation is presented in Table 2.

### 3.2 Data Uncertainty Modelling

Two types of uncertainty were modelled in this application: one relative to the rainfall, and one due to parcel mixture in tanks.

**Rainfall Uncertainty.** The rain is a phenomenon that is geographically very variable, especially in tropical areas. In the plantation, all the rain data come from a single meteorological station located inside the plantation. Since the plantation area is very large, it is sensible to make the hypothesis that the farther is located a parcel in the plantation from the meteorological station, the more uncertain is its rainfall data. This uncertainty is non-probabilistic and progressive, so we chose to model it with a consonant mass function. Moreover, as more rainfall implies more uncertainty, it is logical to assume that imprecision of focal elements increases multiplicatively (i.e. more rainfall measured by the station implies wider focal elements).

To keep the complexity of the uncertainty reasonable, we limited the mass to five focal elements of the form: $[w(1-\delta), w(1+\delta)]$ where $w$ is the original precise rainfall data and where $\delta \in \Delta = \{0, 0.25, 0.5, 0.75, 1\}$. The proposed model is easy to expand to more than five focal elements and can therefore accommodate various levels of complexity (depending on the available computational power and on possible time constraints).

**Table 1.** Variables characteristics

| variable | type | category | unit |
|---|---|---|---|
| period | categorical | agronomical | |
| season | categorical | climatic | |
| weight | numerical | agronomical | kg |
| X (latitude) | numerical | geographical | km |
| Y (longitude) | numerical | geographical | km |
| clone | categorical | agronomical | |
| panel | categorical | agronomical | |
| tapping system | categorical | agronomical | |
| surface | numerical | agronomical | hectar |
| planting year | numerical | agronomical | year |
| first tapping year | numerical | agronomical | year |
| tapping age | numerical | agronomical | year |
| annual number of tapped trees | numerical | agronomical | |
| tapped tree per hectar | numerical | agronomical | |
| temperature | numerical | climatic | celcius degrees |
| relative humidity | numerical | climatic | |
| sun hours | numerical | climatic | hours |
| rainfall | numerical | climatic | mm |
| P30 | numerical | agronomical | |

**Table 2.** $P_{30}$ discretisation

| class labels | $P_{30}$ range |
|---|---|
| very bad | [1.87 ; 14.7[ |
| bad | [14.7 ; 21.6[ |
| medium | [21.6 ; 27.4[ |
| good | [27.4 ; 32.9[ |
| very good | [32.9 ; 49.5[ |

We define a function $g : [0, d_{max}] \times \Delta \rightarrow [0,1]$ ($d_{max}$ being the maximal distance between a location of interest and the measurement station) such that the rainfall $w$ of a parcel located at a distance $d$ from the meteorological is characterized by $g(d,\delta) = m^W([w(1-\delta), w(1+\delta)])$ for all $\delta \in \Delta$.

We distinguish two types of focal elements, the most precise ones ($\delta < 0.5$), and the most imprecise ones ($\delta \geq 0.5$). In the first case (precise ones), we assume that the farther was the parcel from the meteorological station ($d$ increasing), the smaller had to be the masses assigned to those focal elements. In the second, we want to assign bigger masses to the farther parcels. Such assumptions can be translated in the following constraints:

$$\begin{cases} \delta < 0.5 \rightarrow \frac{\partial g}{\partial d} < 0 \\ \delta \geq 0.5 \rightarrow \frac{\partial g}{\partial d} > 0 \end{cases} \tag{2}$$

that merely translate the assumption into derivative behaviors. As function $g$ is used to define mass functions, we must add the following constraints

$$\forall (d,\delta) \in [0,d_{max}] \times \Delta, \quad g(d,\delta) \geq 0 \tag{3}$$

$$\forall d \in [0,d_{max}], \quad \sum_{\delta \in \Delta} g(d,\delta) = 1 \tag{4}$$

that simply ensure one that the obtained mass functions will be well-defined, i.e. that it will be positive (constraint (3)) and will sum up to one (constraint (4)).

One simple solution of this problem is to use two linear functions, one increasing for the most precise focal elements ($\delta < 0.5$) and one decreasing for the most imprecise ones ($\delta \geq 0.5$), and to use a convex sum of those two functions. We obtain:

$$g(d,\delta) = \delta(\frac{2d}{5d_{max}}) + (1-\delta)(\frac{2}{5} - \frac{2d}{5d_{max}}) \tag{5}$$

*Example 2.* Consider three rainfall measurements $w_1 = 0$, $w_2 = 10$ and $w_3 = 30$ from the station, and for each of these measurements some corresponding parcels of interest respectively distant of $20km$, $50km$ and $2km$ from the station. Assuming that $d_{max} = 80$, we obtain $m^{w_1}(\{0\}) = 1$, given the multiplicative uncertainty, and

$$\begin{cases} m^{w_2}(\{10\}) = 0.15 \\ m^{w_2}([7.5,12.5]) = 0.175 \\ m^{w_2}([5,15]) = 0.200 \\ m^{w_2}([2.5,17.5]) = 0.225 \\ m^{w_2}([0,20]) = 0.250 \end{cases} \quad \begin{cases} m^{w_3}(\{30\}) = 0.39 \\ m^{w_3}([22.5,37.5]) = 0.295 \\ m^{w_3}([15,45]) = 0.2 \\ m^{w_3}([7.5,52.5]) = 0.105 \\ m^{w_3}([0,60]) = 0.01. \end{cases}$$

The absence of rain is thus considered certain ($m^{w_1}$) whereas positive rainfall data masses are concentrated on imprecise focal elements when coming from distant parcels ($m^{w_2}$) and on more precise ones when coming from parcels close from the meteorological station ($m^{w_3}$).

**Parcelar Mixtures Uncertainty.** During the harvest, the latex coming from many parcels is usually mixed in some tank. All the parcel variables (i.e., most agronomical variables of Table 1) we have are therefore subject to uncertainty as the amount of latex coming from each parcel in tanks is not tracked. During a pre-treatment of the data, we therefore split all those parcel variables into rough and uncertain proportions (due to latex production high variability) computed from the weight of latex produced annually by each parcel (shifting from 18 original attributes to 106, with all the split ones being in range $[0,1]$).

For example, if 25% of a tank content comes from clone A parcels and 75% from clone B parcels, the actual amount of clones A and B latex in the tank may be quite different, as each clone has variable production capacities (that may depend differently on the weather, soil conditions, etc.). We model this uncertainty such that the more balanced are the parcel proportions in a tank, the more uncertain become those proportions: proportions of a tank with latex from only one pure parcel should remain certain,

while proportions of a tank with equal amounts of latex coming from different parcels should be maximally uncertain.

To do that we used simple intervals around computed crisp proportions, with the interval width increasing as proportions become uniform. To measure this uniformity we simply used the Shannon entropy denoted *ent* computed on the set of parcel proportions of the tanks. The obtained model is the following: for each parcel variable $X^j$ having $r$ modalities with the positive proportions $\{p_1,\dots,p_r\}$,

$$\begin{cases} m^{j_1}([\max(p_1-\frac{ent(p_1,\dots,p_r)}{r},0),\min(p_1+\frac{ent(p_1,\dots,p_r)}{r},1)]) = 1 \\ \vdots \\ m^{j_r}([\max(p_r-\frac{ent(p_1,\dots,p_r)}{r},0),\min(p_r+\frac{ent(p_1,\dots,p_r)}{r},1)]) = 1 \end{cases} \tag{6}$$

with $m^{j_i}$ modelling the uncertainty about the $j_i$th proportion.

*Example 3.* Let us consider a tank containing 75% of clone A and 25% of clone B. The entropy on those proportions is equal to 0.8113. The obtained masses are therefore

$$\begin{cases} m^{clone\,A}([34.43\%,100\%]) = 1 \\ m^{clone\,B}([0\%,65.57\%]) = 1 \end{cases}$$

### 3.3 Experiments

In order to see the consequences of integrating data uncertainty the way we described in Section 3.2, we perform some experiments comparing standard *CART* trees and $E^2M$ trees on the original precise dataset and its corresponding evidential dataset obtained with our uncertainty models, respectively.

For both methodologies, the stopping criteria is a maximum of 5 leaves (to preserve a high interpretability) and a relative purity gain of 0.05.The error rates were computed as the proportion of misclassified examples in the test dataset. Their means were computed from ten 3-fold cross validations. It is noticeable that we used standard (precise error rates) even for the $E^2M$ decision trees for comparison purposes. Given the small tree size, no pruning is done.

**Table 3.** Results

| methodology | mean error rates | 95% confidence interval |
|---|---|---|
| *CART* | 0.6607 | [0.6314 ; 0.6900] |
| $E^2M$ | 0.6560 | [0.6266 ; 0.6854] |

As shown in Table 3, the accuracies of the two methodologies are quite similar, even if the $E^2M$ accuracy is slightly better. Let us now shift to the main interesting part for the experts (and hence for the application goal): the interpretation of the trees.

### 3.4 Interpretation

In a pure knowledge discovery concern, we learnt *CART* and $E^2M$ decision trees on the whole dataset in order to compare the informations they provide about the natural rubber quality explanation. Within such a goal, it should also be noted that integrating data uncertainty makes the result somehow more faithful to our actual information (and therefore more reliable for the experts). The learning parameters were exactly the same as in Section 3.3.

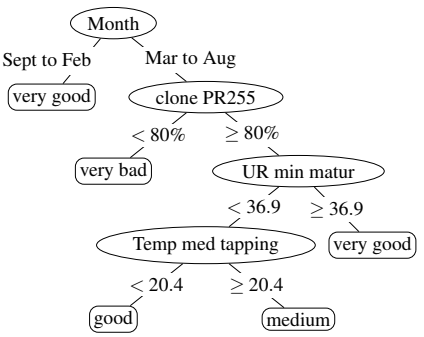



**Fig. 3.** *CART* decision tree

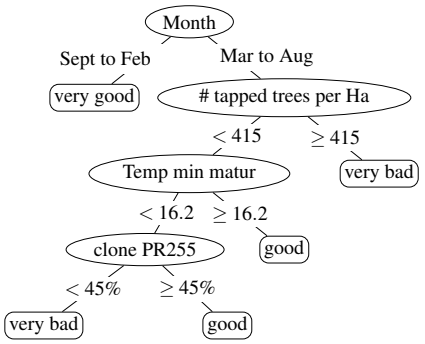



**Fig. 4.** $E^2M$ decision tree

In Figures 3 and 4 that show the learning results, UR min matur is the minimum of relative humidity during the maturation of the latex in buckets, Temp med (resp. min) tapping is the medium (resp. minimum) temperature during the tapping day, # of tapped tree per Ha is the number of tapped trees per hectare.

As shown in those figures, integrating data uncertainty makes the number of tapped trees per hectare appear in the tree. Not only the $E^2M$ decision tree suggests a high relation between quality and productivity, but it also provides a density acceptable bound inside the plantation. According to the experts, this issue points out the potential need to investigate further this path. We can also notice that this data uncertainty lessen the role of clone PR255 proportion in the rubber quality explanation.

## 4 Discussion

In the rubber application, the prediction results from *CART* and $E^2M$ decision trees do not seem very different in term of accuracy (even if $E^2M$ has a slight advantage), but the interpretation of the trees can be quite different; this latter interpretation being as important as accuracy in many applications. Indeed, we observe that some variables may play a significant role in rubber quality explanation once their uncertainty is modelled (here by belief functions). Modelling uncertainty also provides more trustful results for the experts.

Perspectives concerning the application include the modelling of additional uncertainties (in particular with respect to the tapping day) as well as additional studies involving more variables (e.g., tree diseases, soil conditions) or other prediction goal (e.g., quantity of production, which is also important from an economical viewpoint).

Finally, as the uncertainty models we have introduced may be useful in other areas (particularly the distance-based model of rainfall), we plan to generalize them and study in more details their properties.

**Acknowledgments.** This work has been done thanks to CIRAD fundings and LASB laboratory support. We also wish to greatly thank Dr. Patrice Loisel for his great help on the rainfall uncertainty model.

## References

1. Zadeh, L.: Fuzzy sets as a basis for theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
2. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
3. Walley, P.: Statistical reasoning with imprecise probabilities (1991)
4. Kanj, S., Abdallah, F., Denœux, T.: Evidential multi-label classification using the random k-label sets approach. In: Denœux, T., Masson, M.-H. (eds.) Belief Functions: Theory & Appl. AISC, vol. 164, pp. 21–28. Springer, Heidelberg (2012)
5. Sutton-Charani, N., Destercke, S., Denœux, T.: Classification trees based on belief functions. In: Denœux, T., Masson, M.-H. (eds.) Belief Functions: Theory & Appl. AISC, vol. 164, pp. 77–84. Springer, Heidelberg (2012)
6. Janikow, C.Z.: Fuzzy decision trees: issues and methods. IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics 28, 1–14 (1998)
7. Shafer, G.: A mathematical theory of evidence (1976)
8. Sutton-Charani, N., Destercke, S., Denœux, T.: Learning decision trees from uncertain data with an evidential em approach
9. Denœux, T.: Maximum likelihood estimation from uncertain data in the belief function framework. IEEE Trans. on Know. and Data Eng. 25, 119–130 (2011)
10. Périnel, E.: Construire un arbre de discrimination binaire à partir de données imprécises. Revue de Statistique Appliquée 47, 5–30 (1999)

# An Interval Programming Approach for an Operational Transportation Planning Problem

Valeria Borodin[1], Jean Bourtembourg[2], Faicel Hnaien[1], and Nacima Labadie[1]

[1] Laboratory of Industrial Systems Optimization (LOSI),
Charles Delaunay Institute (ICD), University of Technology of Troyes (UTT),
12 rue Marie Curie - CS 42060, 10004 Troyes, France
{valeria.borodin,faicel.hnaien,nacima.labadie}@utt.fr
[2] Agricultural Cooperative Society in the Region of Arcis-sur-Aube,
Industrial Zone of Villette, Villette-sur-Aube 10700, France
j.bourtembourg@scara.fr

**Abstract.** This paper deals with an interval programming approach for an operational transportation problem, arising in a typical agricultural cooperative during the crop harvest time. More specifically, an interval programming model with uncertain coefficients occurred in the right-hand side and the objective function is developed for a single-period multi-trip planning of a heterogeneous fleet of vehicles, while satisfying the stochastic seed storage requests, represented as interval numbers. The proposed single-period interval programming model is conceived and implemented for a real life agricultural cooperative case study.

**Keywords:** interval linear programming, single-period multi-trip transportation planning problem, OR in agriculture.

## 1 Introduction

Gathering the harvest is considered one of the most crucial activity in agricultural environment for both cooperatives and individual farmers, in terms of the high costs involved and the vulnerability to weather conditions. Logistics and transportation activities constitute an inherent and primordial component of the agricultural cooperative system [2,5].

In most cases, farmers deliver their harvest to the nearest storage facility. For the sake of proximity storage availability, the cooperative forwards received cereals from buffer silos to expedition ones. Hence, new crop quantities are received as time progresses, which leads require planning the transfer and transportation of stored seed products.

Satisfaction of farmers' grain storage requests and high level reception service represent the main priority of the agricultural cooperatives during the harvest season. Quantities to be received at each storage facility are ordinarily unknown. In this regard, the case study cooperative has only an approximate information based-on prediction of the daily farmers' crop delivery quantities, represented as interval numbers.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 117–126, 2014.

The purpose of this paper is to present a single-period multi-trip transportation planning application, modelled as a linear programming model with interval objective function and right-hand side constraints. It was motivated by a real case study encountered at a typical French agricultural cooperative.

The remainder of this paper is structured as follows: in the next section, a review of literature related to the interval programming research and applications is provided. The problem statement and modelling is presented in Section 3. After that, in Section 4, the solution method is exposed. In Section 5, several computational experiments are reported and discussed. Finally, in Section 6, overall remarks are drawn and topics for future research are outlined.

## 2 Literature Review

The conjecture of many real life problems presupposes miscellaneous types of uncertainties ([14],[23]). Classical mathematical programming models, nonetheless, only cope with deterministic values of problem input data. With the requirement of tackling non-deterministic data, appropriate techniques have been developed to suit various purposes and for different features of the stochastic data representation: probabilistic, possibilistic, and/or interval formats.

For decision-making problems considering uncertainty, the stochastic linear programming models touch effectively upon various random data with known probability distributions [12,22]. On this topic, the simple recourse model (a type of two-stage or multi-stage stochastic programming) consists in optimizing the expected objective function subject to some expected constraints [20]. Integrated chance-constrained programming is another approach in which solutions are feasible for a discrete set of probabilistic scenarios and all soft constraints are satisfied simultaneously with a given confidence level [8,18]. In turn, dependent chance-constrained programming pertains to maximizing some chance functions of events defined on stochastic sets in a complex uncertain decision system [16,15]. In fuzzy linear programming, the constraints and the objective function are regarded as fuzzy sets and their membership functions are assumed to be known.

Nevertheless, it turns out to be often difficult to specify a relevant membership function or an appropriate probability distribution in an stochastic environment [21,11]. In the last two decades, many theoretical studies have been focused on solving interval linear programming problems, in which the bounds on the uncertain data variation are required, without insisting on their probability distributions or membership functions. The interval analysis method was pioneered by Moore in 1959 as a tool for automatic control of the errors in a computed results [24].

Interval programming models where only the coefficients of the objective function are random and represented in interval format, are studied in [10,1,3,13]. Related to this matter, min-max regret optimization approach is usually employed, where different criteria that can be optimized, are distinguished: worst-case, best-case, worst-case relative regret, worst-case absolute regret and maximum regret

criteria, respectively. In this context two kinds of optimality (i.e. possible/weak and necessary/strong optimality) are defined.

Another stream of recent literature considers the interval left-hand side linear programming. In the work designed by [6], the authors have incorporated stochastic coefficients with multivariate normal distribution within an interval parameter linear programming context. On the other hand, when randomness, represented as interval numbers, appears in right-hand side constraints, only few results have already been obtained. The difficulty resides in the fact that the set of feasible solutions in not exactly known. In [9] the authors have investigated the complexity of two optimization versions, which correspond to the worst and best optimum solutions, when each right-hand side coefficient is defined as an interval number.

However, to the best of our knowledge and as remarked by [17], there are only few results on the issue of optimal solutions for a general interval linear programming, where the vector of the objective function, the coefficient matrix and the right-hand side are all interval matrices.

On the other hand, interval linear programming seems to be a sound approach to face uncertainty issues that are met in real life applications. Nonetheless, it is worth pointing out that a very few number of papers can be found in which interval linear programming applications are developed, notwithstanding its possible wide usage for modelling and solving real world problems [6,7].

## 3 Problem Statement and Modelling

Let us consider an agricultural cooperative specialized in multi-seed production and commercialization. The cooperative involves several hundred of farmers, for whom it provides consulting, drying, storage, transportation and many other customer services.

Once the cereals have reached their physiological maturity, they are harvested and carefully forwarded towards storage facilities (also named *silos*), designed especially for this purpose. Many cooperatives use two types of silos: *expedition silos* $E$, used for a long time period storage, and *buffer silos* $B$, which serve as proximity facilities at the time of harvest. Due to limited storage capacity of buffer silos, an inventory control level and a daily grain transfer to expedition silos are organized during whole harvest season. This ensures the buffer silos availability, which contributes to increase the level of reception and storage services.

Heterogeneous vehicle fleet $K$, previously dimensioned, is dedicated to empty the buffer silos $b$ ($b \in B$), whose cells $c$ ($c \in C_b$) are quickly filling up as harvest time progresses. In order to maintain buffer silos sufficiently empty, a regular (single-period) cereal transfer (delivery) is organized from buffer silos to expedition ones.

More precisely, in each period of time (day) a multi-trip planning is performed in order to guarantee a sufficient silo capacity for receiving the quantities of different varieties $v$ expected to be delivered in the following periods of time.

All of the above transfer activities are realized by seeking to minimize the exceeded quantities to be received in the next $p$ periods of time, that cannot by adequately stored with respect to the current buffer silo $b$ stock level of each cell $c$, $s_{cbv}$.

**Table 1.** Modelling notation

| | |
|---|---|
| **Parameters:** | |
| $v, v \in V$ | crop variety index |
| $b, b \in B$ | buffer silo index |
| $e, e \in E$ | expedition silo index |
| $k, k \in K$ | vehicle index |
| $c, c \in C_b$ | silo cell index |
| $p, p \in P$ | a short time horizon |
| **Deterministic Data:** | |
| $s_{cbv}$ | stock level of variety $v$ in the cell $c$ of buffer silo $b$ at the beginning of the planning period |
| $u_{cb}$ | capacity of cell $c$ of buffer silo $b$ |
| $h_{cbv}$ | 1, if the cell $c$ is allowed to stock the variety $v$ and 0, otherwise |
| $r_{v'v''}$ | 1, if varieties $v'$ and $v''$ are compatible and 0, otherwise |
| $g_k$ | capacity of vehicle $k$ |
| $t_{eb}$ | travel time between the silos $e$ and $b$ |
| $t_l$ | loading time |
| $t_u$ | unloading time |
| $T$ | daily working time |
| $M$ | big value, e.g. greater than or equal to $u_{cb}$ |
| **Stochastic Data:** | |
| $q_{bv}^{\pm}$ | quantity of variety $v$ expected to be delivered to buffer silo $b$ during the period of time ahead |
| $\tilde{q}_{bv}^{\pm}$ | quantity of variety $v$ expected to be delivered to buffer silo $b$ in $p$ following periods of time |
| **Variables:** | |
| $y_{cb}$ | quantity of cell $c$ of buffer silo $b$ to be transferred |
| $z_{cbv}$ | 1, if the cell $c$ of buffer silo $b$ contains the variety $v$ |
| $w_{cbv}$ | available capacity for stocking the variety $v$ of the cell $c$ of buffer silo $b$ |
| $x_{cbe}^{k}$ | number of round trips between the silos $b$ and $e$ for emptying the cell $c$ of silo $b$ |
| $f_{vb}$ | exceeded quantity of variety $v$ at silo $b$, in terms of the $\tilde{q}_{bv}^{\pm}$ against the current silo stock level |

The daily quantities to be received at each storage facility are unknown. In this sense, the case study cooperative has only an approximate information based-on statistical analysis and prediction of the daily farmers' crop delivery quantities, valued as interval numbers, whilst considering the meteorological repercussion on the progress and achievement of the gathering activity and the farmers' delivery behaviour. Therefore, let us denote by $q_{bv}^{\pm}$ the quantity of variety $v$ to be delivered

to the buffer silo $b$ in the next period of time. The uncertain data $\tilde{q}_{bv}^{\pm}$ are defined to represent the forecasted quantity interval of the variety $v$ to be received by the buffer silo $b$ in $p$ following periods of time.

Before proceeding to the problem modelling, let us consider moreover the following problem assumptions:

- Each vehicle can transport per trip goods belonging only to one buffer cell.
- The vehicles start from the first expedition silo (depot point) foreseen in their respective daily multi-trip planning. Respectively, they return to the last expedition silo (ending point) due in its daily multi-trip planning.
- Speed of vehicle is given and fixed. No traffic jam is considered.
- The total working time of each vehicle is limited to $T$ per day.
- Non-Euclidean distance is considered between any two transporting points. Thus, the travel time $t_{eb}$, from silos $e$ to $b$, is not equal to $t_{be}$, from $b$ to $e$.
- For all buffer and expedition silos, the loading $t_l$ and unloading $t_u$ times are given and fixed.

The decision integer variables $x_{cbe}^{k}$ denote the number of round trips made by the vehicle $k$ between the buffer silo $b$ and the expedition silo $e$, for emptying a quantity $y_{cb}$ from the cell $c$ of buffer silo $b$. By the same token, the decision variables $w_{cbv}$ represent the available capacity of silo $c$ of buffer silo $b$ for receiving the variety $v$, while respecting their compatibility $h_{cv}$ and the total cell capacity $u_{cb}$. The data $h_{cv}$ are defined to take the value 1, if the cell $c$ is allowed to stock the variety $v$ and 0, otherwise. In this manner, the cereal allotment pursuing and the variety-cell management are considered. Additionally, an inter-varietal compatibility $r_{v'v''}$ must also be taken into account for a suitable seed nature allotment and traceability. The data $r_{v'v''}$ take the value 1, if varieties $v'$ and $v''$ are compatible and the value 0, otherwise. Two varieties are considered compatible, if they can be mixed and stored in the same cell.

The decision positive variables $f_{vb}$ express the exceeded quantities of each variety $v$, in terms of expected quantity $\tilde{q}_{bv}^{\pm}$ to be delivered in the following $p$ periods of time to the buffer silo $b$, with reference to the total available silo storage capacity of variety $v$, $\sum_{c \in C_b} w_{cbv}$.

In order to guarantee an appropriate storage service, buffer silos must be emptied in such a way to minimize the exceeded quantities at each buffer silos in the following $p$ periods of time, in terms of each expected variety to be delivered and its associated quantity. Subsequently, by considering the defined decision variables and data parameters introduced above (see Table 1), a linear programming model with interval right-hand sides and objective function (1)-(15) is formalized hereafter:

$$min \quad \sum_{v \in V} \sum_{b \in B} f_{vb} \tag{1}$$

subject to:

$$f_{vb} \geq \tilde{q}^{\pm}_{bv} - \sum_{c \in C_b} w_{cbv} \qquad \forall v, \forall b \qquad (2)$$

$$\sum_{c \in C_b} w_{cbv} \geq q^{\pm}_{bv} \qquad \forall v, \forall b \qquad (3)$$

$$u_{cb} - \sum_{v \in V} s_{cbv} + y_{cb} = \sum_{v \in V} w_{cbv} \qquad \forall b, \forall c \qquad (4)$$

$$w_{cbv} \leq M \cdot h_{cbv} \qquad \forall v, \forall b, \forall c \qquad (5)$$

$$1 + w_{cbv} > z_{cbv} \qquad \forall v, \forall b, \forall c \qquad (6)$$

$$w_{cbv} \leq M \cdot z_{cbv} \qquad \forall v, \forall b, \forall c \qquad (7)$$

$$z_{cbv'} + z_{cbv''} \leq r_{v'v''} + 1 \qquad \forall v', v''(v' \neq v''), \forall b, \forall c \qquad (8)$$

$$\sum_{k \in K} \sum_{e \in E} x^k_{cbe} \cdot g_k \geq y_{cb} \qquad \forall c, \forall b \qquad (9)$$

$$\sum_{c \in C_b} \sum_{b \in B} \sum_{e \in E} (t_{eb} + t_{be} + t_l + t_u) \cdot x^k_{cbe} \leq T \qquad \forall k \qquad (10)$$

$$w_{cbv} \geq 0 \qquad \forall c, \forall b, \forall v \qquad (11)$$

$$z_{cbv} \in \{0, 1\} \qquad \forall c, \forall b, \forall v \qquad (12)$$

$$y_{cb} \geq 0 \qquad \forall b, \forall c, \forall k \qquad (13)$$

$$x^k_{cbe} \in \mathbb{N} \qquad \forall e, \forall b, \forall c, \forall k \qquad (14)$$

$$f_{vb} \geq 0 \qquad \forall v, \forall b \qquad (15)$$

Exceeded quantity of each variety at each buffer silo is calculated by the constraints (2), in terms of the expected quantity to be received in the following $p$ periods of time against the current silo stock level. Constraints (3) ensure an available silo capacity for stocking the quantity for each seed variety foreseen to be delivered in the following time period. Stock equilibrium constraints for each silo cell are expressed by (4). Constraints (5) verify if the cells $c$ of buffer silo $b$ is allowed to stock the varieties $v$. Constraints (6) and (7) impose the binary variable $z_{cbv}$ to take the value 1, if the cell $c$ of buffer silos $b$ is reserved to stock the variety $v$ in the next period of time. This is prescribed for respecting the inter-varietal compatibility in each cell $c$ of buffer silo $b$, required by constraints (8). In order to guarantee a sufficient available capacity of the buffer silo, the constraints (9) trigger a seed transfer from buffer silos $b$ to expedition ones $e$, which is performed by using a heterogeneous vehicle fleet $K$. Constraints (10) confine to $T$ the total working time of each vehicle $k$. The objective function (1) seeks to minimize the exceeded quantity expected to be received in the following $p$ periods of time against the current silo stock level.

In the interval programming model (1)-(15), uncertainty, represented by intervals, concerns both the objective function and the right-hand side constraints. Hence, the set of feasible solutions is not exactly known and any solution may be

not feasible for all interval right-hand side constraints. Correspondingly, classical min-max optimization criteria cannot be directly employed [9].

## 4 Solution Methods

An interval programming model (1)-(15) with interval coefficients, simultaneously occurring in the objective function and right-hand side constraints are considered. In this context, the aim consists of determining the best possible optimum and the worst one over all possible configurations, which correspond to an assignment of plausible values for each of the model uncertain parameters.

As far as uncertainty on objective function coefficients is regarded, two criteria are classically considered: the worst case criterion and the best case one. Let $X$ be the set of (1)-(15) problem feasible solutions. Given $x \in X$, the configuration to be considered is the one that corresponds to the worst (best) for this solution. In this sense, the value of $x$, noted $f_{\text{worst}}(x)$ ($f_{\text{best}}(x)$) is defined as presented below:

$$f_{\text{worst}}(x) = \max_{\tilde{q}^- \leq \tilde{q} \leq \tilde{q}^+} \sum_{v \in V} \sum_{b \in B} f_{vb} \tag{16}$$

$$f_{\text{best}}(x) = \min_{\tilde{q}^- \leq \tilde{q} \leq \tilde{q}^+} \sum_{v \in V} \sum_{b \in B} f_{vb} \tag{17}$$

where $\tilde{q}^{\pm} = (\tilde{q}^{\pm}_{bv})_{b \in B, v \in V}$.

The problem is to determine the solution $x_{\text{worst}}(x_{\text{best}})$, which minimizes $f_{\text{worst}}(x)$ and $f_{\text{best}}(x)$ respectively, as follows:

$$f_{\text{worst}}(x_{\text{worst}}) = \min_{x \in X} f_{\text{worst}}(x) \tag{18}$$

$$f_{\text{best}}(x_{\text{best}}) = \min_{x \in X} f_{\text{best}}(x) \tag{19}$$

On the other hand, classical criteria cannot be directly applied when uncertainty concerns right-hand side constraints. Denote by $P^q$ the program (1)-(15), where $q$ varies in the interval $q^- \leq q \leq q^+$, $q^{\pm} = (q^{\pm}_{bv})_{b \in B, v \in V}$. In the context of linear programs with interval right-hand sides, the objective of the best (worst) optimal solution problem is to determine the minimum (maximum) value $\vartheta(P^q)$ of the optimal solution of $P^q$, when $q$ varies in the interval $[q^-, q^+]$. Let us formalize the best optimal solution problem (BEST) and the worst optimal solution problem (WORST), hereafter:

$$\text{BEST} : \begin{cases} \min\ \vartheta(P^q) \\ \text{s.t } q^- \leq q \leq q^+ \end{cases} \tag{20}$$

$$\text{WORST} : \begin{cases} \max\ \vartheta(P^q) \\ \text{s.t } q^- \leq q \leq q^+ \end{cases} \tag{21}$$

Let $X^{\text{BEST}}$ ($X^{\text{WORST}}$) be the set of optimal solutions of BEST (WORST). In the light of the above mentioned approaches, four cases are studied in this paper, for handling the interval programming model (1)-(15):

$$f_{\text{worst}}^{\text{WORST}} = \min_{x \in X^{\text{WORST}}} f_{\text{worst}}(x) \tag{22}$$

$$f_{\text{worst}}^{\text{BEST}} = \min_{x \in X^{\text{BEST}}} f_{\text{worst}}(x) \tag{23}$$

$$f_{\text{best}}^{\text{WORST}} = \min_{x \in X^{\text{WORST}}} f_{\text{best}}(x) \tag{24}$$

$$f_{\text{best}}^{\text{BEST}} = \min_{x \in X^{\text{BEST}}} f_{\text{best}}(x) \tag{25}$$

Criteria (22), (23), (24) and (25) allow to provide the best possible optimum and the worst one over all possible configurations in order to reveal a kind of robustness information, by handling both uncertainty on objective function and right-hand sides. As [19] stated, the range of the objective function between the best and the worst optimum values provides an overview of the risk involved, which can be reflected by specifying the values of the uncertain coefficients.

## 5 Computational Results

The model (1)-(15) was conceived to better organize the buffer silos emptying during the harvest season for an Agricultural Cooperative Society, situated in the region of Arcis-sur-Aube (France). More precisely, this is a grain and oilseed agricultural cooperative, for which an appropriate buffer silos emptying policy and multi-trip transportation planning is necessary to ensure a high level seed reception and storage services.

The interval programming model (1)-(15) has been implemented by using the `C++` language. Criteria (22), (23), (24) and (25) have been used and the corresponding models have been solved by using `ILOG CPLEX` (version 12.6.0) optimization software, for instances with up to 30 varieties, 3 expedition silos, 11 buffer silos and between 10 and 15 cells per buffer silo (whose capacities vary between 200 and 800 tonnes per cell). Computational experiments have been carried out on an Intel(R) Core(TM) i7-2720QM CPU 2.20GHz workstation.

Commonly, the harvest lasts about one month. During this period, the expected quantities to be received by each case study silo were estimated based-on cooperative predictive modelling of the farmers' crop delivery behaviour and climate forecasting data, derived from nearby weather stations with an acceptable reliability level. Due to a small gap between varieties' ripeness dates and to a high farmers' gathering yield, the value of $p$ was empirically fixed to 3.

In what follows, let us examine the figures Fig. 1 et Fig. 2, which illustrate the output results corresponding to the best possible optimum and the worst optimum over all possible configurations for a time horizon of 7 harvest days. More specifically, the figure Fig. 1 reports the optimal values for (22) and (23), as well as, the figure Fig. 2 provides the optimal value of (24) and (25), respectively.

Representative gaps between the objective values of approaches (22)-(23) and (24)-(25) corresponding to the periods 2, 3 and 6, suggest about eventual considerable buffer storage unavailability. It could be due to the fact that a significant range of varieties is expected to be delivered. For preventing unsuitable seed nature allotment or quality degradation, the cooperative should rent supplementary vehicle during the respective periods of time. Contrarily, the multi-trip

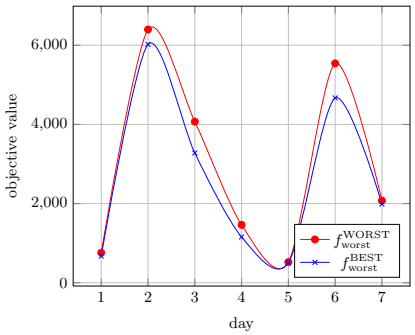



**Fig. 1.** (22) vs. (23)

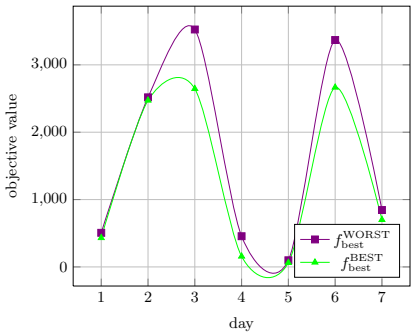



**Fig. 2.** (24) vs. (25)

planning solutions provided by (22) could be realised during the periods 1, 4 and 5, since negligible objective values and gaps are recorded for these periods.

As computational results pointed out, approaches (22), (23), (24), (25) help to handle efficiently the inventory control and multi-trip transportation planning problem by presenting good alternative solutions. They offer a pertinent decision support by taking into account weather and farmers' delivery uncertainties.

## 6 Conclusions and Topics for Future Research

This paper presents an interval programming model for a single-period multiple trip transportation planning problem, for purpose of maintaining available cooperative buffer silos during the harvest season. Best and worst optimum criteria, prescribed to deal with uncertainty on objective function, have been considered for both best and worst optimal solution problems, which address uncertainty on right-hand side coefficients.

Future research would be dedicated to tackle and study other approaches of problem robustness (e.g. maximum regret criterion, etc.). Moreover, other problem formulations would be also tested to deal with the problem considered in this paper (e.g. composing the rented fleet of vehicles, whilst ensuring the buffer silos availability, etc.).

## References

1. Aissi, H., Bazgan, C., Vanderpooten, D.: Min-max and min-max regret versions of combinatorial optimization problems: A survey. European Journal of Operational Research 197(2), 427–438 (2009)
2. Ahumada, O., Villalobos, J.: Application of planning models in the agri-food s upply chain: A review. European Journal of Operational Research 196(1), 1–20 (2009)
3. Averbakh, I.: Computing and minimizing the relative regret in combinatorial optimization with interval data. Discrete Optimization 2(4), 273–287 (2005)

4. Bertsimas, D., Sim, M.: The price of robustness. Operations Research 52(1), 35–53 (2004)
5. Borodin, V., Bourtembourg, J., Hnaien, F., Labadie, N.: A discrete event simulation model for harvest operations under stochastic conditions. In: ICNSC 2013, pp. 708–713 (2013)
6. Cao, M., Huang, G., He, L.: An approach to interval programming problems with left-hand-side stochastic coefficients: An application to environmental decisions analysis. Expert Systems with Applications 38(9), 11538–11546 (2011)
7. Dai, C., Li, Y., Huang, G.: An interval-parameter chance-constrained dynamic programming approach for capacity planning under uncertainty. Resources, Conservation and Recycling 62, 37–50 (2012)
8. Dentcheva, D., Prékopa, A., Ruszczyński, A.: Bounds for probabilistic integer programming problems. Discrete Applied Mathematics, 55–65 (October 2002)
9. Gabrel, V., Murat, C., Remli, N.: Linear programming with interval right hand sides. International Transactions in Operational Research 17(3), 397–408 (2010)
10. Inuiguchi, M., Sakawa, M.: Minimax regret solution to linear programming problems with an interval objective function. European Journal of Operational Research 86(3), 526–536 (1995)
11. Kacprzyk, J., Esogbue, A.: Fuzzy dynamic programming: Main developments and applications. Fuzzy Sets and Systems 81(1), 31–45 (1996)
12. Kall, P., Wallace, S.N.: Stochastic Programming. John Wiley and Sons (1994)
13. Kasperski, A., Zieliński, P.: Minmax regret approach and optimality evaluation in combinatorial optimization problems with interval and fuzzy weights. European Journal of Operational Research 200(3), 680–687 (2010)
14. List, G., Wood, B., Nozick, L., Turnquist, M., Jones, D., Kjeldgaard, E., Lawton, C.: Robust optimization for fleet planning under uncertainty. Transportation Research Part E: Logistics and Transportation Review 39(3), 209–227 (2003)
15. Liu, B.: Dependent-chance programming: A class of stochastic optimization. Computers & Mathematics with Applications 34(12), 89–104 (1997)
16. Liu, B., Iwamura, K.: Modelling stochastic decision systems using dependent-chance programming. European Journal of Operational Research 101(1), 193–203 (1997)
17. Luo, J., Li, W.: Strong optimal solutions of interval linear programming. Linear Algebra and its Applications 439(8), 2479–2493 (2013)
18. Miller, L., Wagner, H.: Chance-constrained programming with joint constraints. Operational Research, 930–945 (1965)
19. Chinneck, J.W., Ramadan, K.: Linear programming with interval coefficients. The Journal of the Operational Research, 209–220 (2000)
20. Ruszczyński, A., Shapiro, A.: Stochastic programming models. Handbooks in Operations Research and Management Science, pp. 1–64. Elsevier (2003)
21. Sengupta, A., Pal, T., Chakraborty, D.: Interpretation of inequality constraints involving interval coefficients and a solution to interval linear programming. Fuzzy Sets and Systems 119(1), 129–138 (2001)
22. Shapiro, A., Dentcheva, D.: Lectures on stochastic programming: Modelling and Theory. Society for Industrial and Applied Mathematics, Philadelphia (2009)
23. Sheu, J.-B.: A novel dynamic resource allocation model for demand-responsive city logistics distribution operations. Transportation Research Part E: Logistics and Transportation Review 42(6), 445–472 (2006)
24. Moore, R.E.: Automatic Error Analysis in Digital Computation. Sunnyvale, Calif. (1959)

# Fuzzy Modeling of a Composite Agronomical Feature Using FisPro: The Case of Vine Vigor

Cécile Coulon-Leroy[1,3], Brigitte Charnomordic[2], Marie Thiollet-Scholtus[3], and Serge Guillaume[4]

[1] LUNAM Université, Groupe ESA, UPSP GRAPPE, Angers, France
c.coulon@groupe-esa.com
[2] INRA, SupAgro, UMR MISTEA, Montpellier, France
bch@supagro.inra.fr
[3] INRA, SAD-UEVV-1117, Colmar, France
marie.scholtus@angers.inra.fr
[4] Irstea, UMR ITAP, Montpellier, France
serge.guillaume@irstea.fr

**Abstract.** Fuzzy logic is a powerful interface between linguistic and numerical spaces. It allows the design of transparent models based upon linguistic rules. The *FisPro* open source software includes learning algorithms as well as a friendly java interface. In this paper, it is used to model a composite agronomical feature, the vine vigor. The system behavior is characterized by its numerical accuracy and analyzed according to the induced knowledge. Well known input output relationships are identified, but also some rules reflect local interactions.

**Keywords:** Fuzzy rules, Learning, Agronomy, Vine

## 1 Introduction

In many application fields, there is a need for interactive computerized system that gather data and knowledge from a wide range of sources, in order to help understanding a complex phenomenon and making a decision. In particular, the application fields of agronomy and environment are in demand for advanced modeling and decision support tools while presenting some specificities. Observational and modeling studies call for various spatio-temporal scales: plot, farm, catchment basin, designation of origin, and many influential parameters have to be considered. Data are acquired with very different means and resolutions, ranging from manual costly measurements such as soil analysis or plant hydric potential to high resolution data from embedded sensors and aerial images. In any case, the variability that characterizes life sciences is pervasive and results in data uncertainty and a lack of reproducibility. Therefore, human expertise is fundamental in interpreting data, and stakeholders in agronomy and environmental have always relied on expertise to interpret observations and to take decisions.

When expertise and data have to be integrated in a reasoning framework, fuzzy logic and Fuzzy Inference Systems (FIS) can play an original part in the

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 127–137, 2014.

modeling and decision support process [1]. Let us take as an example the case of vine growing, that is a complex agricultural system involving several factors. The vegetative vine development is called vigor. It takes into account the rhythm and the intensity of the vine shoot growth. Empirically, in relative terms, vine vigor level of a plot is well known as being stable over the years. It is highly influenced by environmental factors, such as soil or climate, but can also be modified by agricultural practices (choice of rootstock, inter-row management, pruning type, among others). Vine vigor is a key parameter to control the balance between vegetative growth and productivity, which influences berry composition and then wine characteristics.

For a composite feature such as vine vigor, it is unrealistic to design formal mathematical models based on ecophysiological knowledge. An alternative approach consists in deriving empirical models from experiments. However, for perennial crops such as grapevine, full experimental designs to test a large number of factors in interaction are very difficult to implement. Furthermore the collected data are tainted with uncertainty; the features can suffer from imprecision, as many assessments are made by human beings. The learning process must be adapted to deal with partial and imperfect data, and to include valuable pieces of expert knowledge.

Various learning methods can be used to produce a model to study interactions between variables. They include artificial intelligence or statistical techniques. Both can deal with some kinds of data imperfection and both have been used in agri-environmental modeling. Common choices include classical linear models and decision trees [2] or, for more recent developments, Bayesian networks [3]. These statistical models are efficient in a wide range of situations, and often yield a confidence interval, since they are based on probability theory. However, they may be difficult to interpret or to use in cases where data imperfection and uncertainty is prevalent. Fuzzy modeling and FIS offer an interesting alternative in such a case, mainly because they provide an interface between the numerical and the linguistic spaces [4].

The objective of the present paper is to show the interest of FIS to study the interactions between environmental factors, agricultural practices and vine vigor. The approach attempts to make the best of domain expertise and of available field data, though they are incomplete, in order to design an interpretable model. The interpretability makes it possible to analyze the system behavior and to evaluate interactions between variables.

Software with a friendly interface is indispensable to allow interactive modeling and decision support. *FisPro*[1] is an open source software that has been recently used in several agronomic and environmental applications. These applications cover different topics: agricultural management using decision variables defined at catchment scale [5]; modeling interactions among sustainability components of an agro-ecosystem [6]; determining optimum rates of nitrogen for corn on the basis of field and crop features [7]; predicting vine vigor and

[1] `http://www.fispro.org/, email: fispro@supagro.inra.fr`

precocity [8]; characterizing management zones from viticulture data [9]; a soil occupation diagnostic for sugarcane harvest [10].

The paper is organized as follows. Section 2 presents some general functionalities for fuzzy modeling and learning with *FisPro*. Section 3 describes the application to vine vigor, and analyzes the results from a knowledge discovery point of view. Some conclusions are given in Section 4.

## 2 FisPro: Fuzzy Inference System Design and Optimization

*FisPro* implements an applied research work in the field of interpretable FIS [11,4,1]. *FisPro* has been used to design FIS in different domains, including economics, hydraulics, robotics, agri-food industry, medicine, agronomy and environment... Among fuzzy software products, *FisPro* stands out because of the interpretability of fuzzy systems automatically learned from data. Interpretability is guaranteed at each step of the FIS design with *FisPro*: variable partitioning, rule induction, optimization.

### 2.1 A Framework for Fuzzy Modeling

A typical FIS is represented in Figure 1. It consists of three stages: 1) fuzzification to transform numerical values into membership degrees in the fuzzy sets associated to linguistic concepts, 2) fuzzy rule base, 3) defuzzification process, to infer a crisp value from the rule aggregation result.

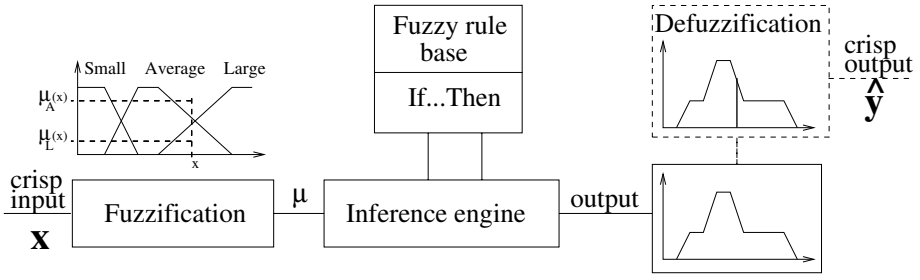



**Fig. 1.** A fuzzy inference system

In the present work, *FisPro* has been used to design FIS, according to the approach summarized in Figure 2.

That approach combines expertise and data. For instance, the number of fuzzy sets is chosen according to expert knowledge, but data make easier the lengthy task of linguistic modeling, by learning fuzzy set characteristic points. Automatic rule learning is done to highlight the interactions that arise from the

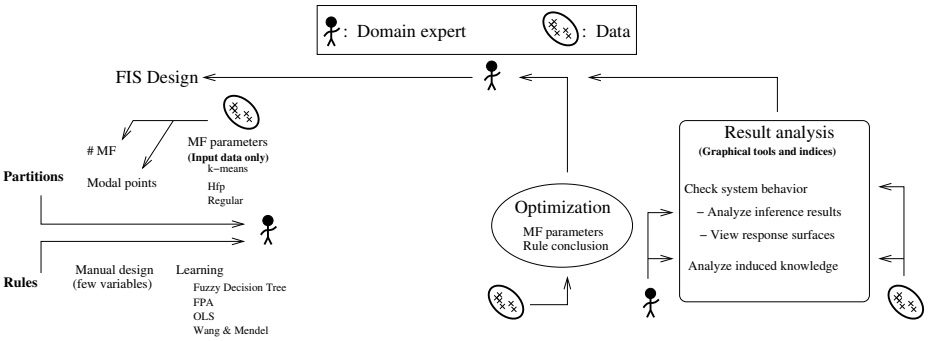



**Fig. 2.** FIS and linguistic modeling with *FisPro*

multidimensional characteristics. Optimization allows FIS parameter tuning. In a case study when data are scarce, as for the study of vine vigor, the FIS output is analyzed from a knowledge discovery point of view, instead of the classical validation step only based on numerical accuracy.

### 2.2 Linguistic Variable and Fuzzy Partitioning with FisPro

Working with the membership degrees in the different linguistic concepts, instead of the raw data values, reduces the system sensitivity to raw data variation. This is a convenient and meaningful way to tackle biological variability.

The readability of fuzzy partitioning is a pre-requisite condition to build an interpretable rule base. The necessary conditions for interpretable fuzzy partitions have been studied by several authors [12]: Distinguishable concepts, a justifiable number of fuzzy sets, coverage (each data point, $x$, should belong significantly, $\mu(x) > \epsilon$, at least to one fuzzy set), normal and significantly overlapping fuzzy sets. These requirements are all fulfilled by the strong fuzzy partitions, illustrated in Figure 3. For each point in the universe, the sum of the membership degrees in all the fuzzy sets is equal to one. Even if, in interactive design, other membership function (MF) shapes are available and fuzzy partitions can be freely adjusted, *FisPro* automatic procedures systematically generate strong fuzzy partitions with semi-trapezoidal shapes at the edges and either triangular or trapezoidal shaped MFs elsewhere.

The process of partitioning comes to choose the number of fuzzy sets and the corresponding characteristic points. When possible, the number of fuzzy sets is determined by expertise, in order to facilitate the interpretation. Variables can be continuous or discrete, under the condition that their values are ordered and have a progressive semantic meaning. Discrete variables are described by k ordered values. The characteristic points of MFs for continuous inputs are not easy to determine only by expertise so a learning procedure can be run (see Figure 2), for instance with the monodimensional k-means algorithm on the

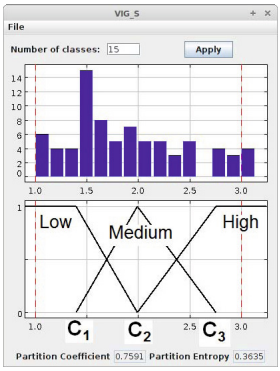



**Fig. 3.** A strong fuzzy partition with three linguistic labels

input data, independently for each variable. Then the cluster centers are chosen as characteristic points.

In *Fispro*, MF partitions can be visualized together with the data distribution, as shown in Figure 3. They can be edited and all modifications are dynamically passed on to other windows, including the inference window.

Several indices have been defined to characterize fuzzy partitions, and are available in *FisPro*. The partition coefficient (PC) and the partition entropy (PE), both proposed by Bezdek [13], are implemented in *FisPro*. Let $N$ be the data set size, $c$ the number of fuzzy sets and $\mu_i(k)$ the membership degree of the $kth$ item in the $ith$ group, the available indices are $PC = \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{c}\mu_i^2(k)$ and $PE = -\frac{1}{N}\left\{\sum_{k=1}^{N}\sum_{i=1}^{c}\left[\mu_i(k)\log_a(\mu_i(k))\right]\right\}$.

According to these criteria a good partition should minimize entropy and maximize the coefficient partition.

### 2.3 Rule Learning and Optimization with FisPro

The fuzzy rules are defined as:

$$IF\ X_1\ is\ A_1^r\ AND\ X_2\ is\ A_2^r\ \ldots AND\ X_p\ is\ A_p^r\ THEN\ y\ is\ C_r\ .$$

**Rule Learning**

It is difficult to define *a best method* for fuzzy rule learning. Several methods are available in *FisPro*, and are detailed in [4]. They all respect the interpretability of the fuzzy partitions. If the number $p$ of input variables is high, it is advisable to select the most influential ones prior to learning, for instance by running fuzzy decision trees. If $p$ is not too high, *WM* (Wang & Mendel) can be used to handle classification cases. For regression cases, when the number $n$ of available data items is high, a statistical inspired method (*Fuzzy Orthogonal Least Squares*) is

interesting to detect outliers and select the most important sources of output variability. The *FPA* method (*Fast Prototyping Algorithm*), which is used in the present work, yields a data summary under the form of fuzzy rules.

FPA principles are recalled below. In a first step, the rules corresponding to the input combinations are generated, only if there are corresponding data in the data set. In a second step their conclusions are initialized according to the data values: $C_r = \frac{\sum_{i \in E_r} W_r(x_i) * y_i}{\sum_{i \in E_r} W_r(x_i)}$, where $W_r(x_i)$ is the matching degree of the ith example for the rth rule, defined as: $W_r(x) = \mu_{A_1^r}(x_1) \wedge \mu_{A_2^r}(x_2) \wedge \ldots \wedge \mu_{A_p^r}(x_p)$ and $E_r$ is a subset of examples chosen according to their matching degree to the rule. If there are not enough items that fire the rth rule with a degree higher than the user defined threshold, the rule is not kept. Thus, FPA yields a subset of all the possible rules. We set the threshold to a membership degree of 0.2, and the minimum cardinality of $E_r$ to 1.

**Optimization and Median FIS**

Parameter optimization allows to optimize all parts of a FIS, using the Solis and Wets algorithm, see [14] for details. As partition parameters and rules have been generated separately, it is interesting to run an optimization procedure of the model as a whole. The optimization algorithm used in this work has been proposed in [1]. It is adapted from Glorennec [15] and based upon the work of Solis and Wets [16]. It allows optimizing all of the FIS parameters: input or output partitions and rule conclusions.

In our approach, the input variables were optimized each in turn, the order depending on the variable importance. To assess that importance, the variables were ranked according to a fuzzy decision tree. The data set was split into a learning set (70% of the vine plots) and a test set (30% of the vine plots). Ten pairs of learning and test sets were randomly created, taking into account the output distribution levels. The optimization procedure was guided by the root mean square error $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \|\widehat{y_i} - y_i\|^2}$, where $\widehat{y_i}$ is the inferred value for the ith item, $y_i$ its observed value and N the number of items. The usual R-squared (R2) was also used to characterize the system accuracy.

The optimization process does not change the system structure; the number of MFs remains the same for all the variables as well as the rule premise structure. Only the MF parameters and the rule conclusions are modified. This allows the semantic properties of the initial model to be preserved while the model accuracy is improved. This also allows to define a *median FIS* from the ten-fold optimization sequence. The *median FIS* fuzzy set characteristic points are the median values of the corresponding values obtained for each fold, as well as the consequent parts of the rules.

*FisPro* provides some useful tools for exploratory analysis of the FIS behavior and representativeness, such as response surfaces and the summary of links

between rules and activated examples. Sampling procedures for cross-validation are integrated in the learning and optimization processes.

## 3 System Analysis

The case study is located in the middle Loire Valley, on the Saumur Protected Designation of Origin (PDO, in French Appellation d'Origine Contrôlée) area, in France. It includes 78 vine plots. All the vineyards considered in the study were planted with *Vitis vinifera* cultivar *Cabernet franc*, the main grape variety in the region, and with similar planting densities. In the studied area, according to expert knowledge, the vine vigor is influenced by soil factors and by two main agricultural practices: rootstock choice and inter-row management. The system was built using three input variables corresponding to the three main influential factors:

1. Vine vigor imparted by soil ($VIG_S$). This indicator is calculated using a fuzzy inference system [8], considering three input variables: the water holding capacity, the gravel percentage in the soil profile and the parent rock hardness. $VIG_S$ is a numerical variable between 1 (Low imparted vigor) and 3 (High imparted vigor) .
2. Vigor conferred by rootstock ($VIG_R$). Vine is grafted on a rootstock to fight against the attack of an insect called *Phylloxera vastatrix*. The rootstock, at the interface between soil and vine variety, interacts with the variety to modify the development of the whole plant. For each rootstock, vigor level was determined from the literature. $VIG_R$ is a discrete variable with five values (1 - Very low; 1.5 - Low; 2 - Medium; 2.5 - High and 3 - Very High).
3. The inter-row management constraint on vine vigor ($VIG_C$). A grass cover is introduced in the vineyard inter-rows to limit runoff and soil erosion. However it also limits vine vegetative development on account of competitions for soil water and nitrogen. $VIG_C$ was defined as a discrete variable with 10 values (between 0 - No constraint and 3 - High constraint). Constraint values were obtained by crossing the constraint imparted by the cover crop variety, thanks to advisory services technical reports, and the cover crop area.

The system output is the vine vigor given by expert assessment, $VIG_{OBS}$. Vigor is linked to the shoot growth and leaf areas observed on vine plots. $VIG_{OBS}$ can take one of the following labels: 1 - Very Low; 2  Low; 3 - High and 4 - Very High. The Medium label was not used on purpose to avoid *safe haven* assessments.

The number of fuzzy sets was determined by expertise, in order to have a number of concepts corresponding to the usual expert vocabulary used by domain experts and technicians. The discrete variable, $VIG_R$, was described by five ordered values. $VIG_S$ and $VIG_C$ were partitioned into three fuzzy sets corresponding to the usual terms Low, Medium and High. The initial fuzzy set characteristics points are given in Table 1 (left).

The rule base learnt by running the FPA procedure described in Section 2 is shown in Table 2 and some comments are given in the following paragraphs.

**Table 1.** The fuzzy partition parameters (left) and FIS accuracy (right)

| | $VIG_S$ | | $VIG_C$ | |
|---|---|---|---|---|
| Param. | Initial | Optimized | Initial | Optimized |
| C1 | 1.4 | 1.4 | 1.00 | 1.02 |
| C2 | 2.0 | 2.1 | 1.65 | 1.50 |
| C3 | 2.8 | 2.1 | 2.25 | 2.18 |

| | FIS | RMSE | $R^2$ |
|---|---|---|---|
| Learning set | Initial | 0.67 | 0.64 |
| | Optimized | 0.52 | 0.77 |
| | Rel. gain | 22% | 20% |
| Test set | Initial | 0.67 | 0.60 |
| | Optimized | 0.54 | 0.73 |
| | Rel. gain | 19% | 22% |

**Table 2.** The fuzzy rule base

| | System inputs | | | Rule conclusions | |
|---|---|---|---|---|---|
| Rules | $VIG_S$ | $VIG_R$ | $VIG_C$ | Initial | Optimized |
| 1 | Medium | Low | High | 2.6 | 2.1 |
| 2 | Low | Medium | Medium | 3.7 | 4.0 |
| 3 | Low | Low | High | 1.3 | 1.2 |
| 4 | Low | Low | Medium | 1.2 | 1.3 |
| 5 | Medium | Low | Medium | 2.5 | 2.4 |
| 6 | Low | Medium | High | 3.5 | 3.8 |
| 7 | High | Low | High | 4.0 | 4.0 |
| 8 | Medium | Medium | Medium | 2.7 | 1.4 |
| 9 | Medium | Medium | High | 2.7 | 1.1 |
| 10 | Low | Medium | Low | 2.5 | 2.2 |
| 11 | High | Medium | Medium | 3.0 | 2.9 |
| 12 | Medium | Medium | Low | 3.3 | 3.2 |
| 13 | High | Medium | High | 3.0 | 2.9 |
| 14 | High | Low | Medium | 4.0 | 4.0 |
| 15 | Low | Low | Low | 3.2 | 3.8 |
| 16 | Low | High | Medium | 4.0 | 3.9 |
| 17 | High | Low | Low | 4.0 | 3.9 |
| 18 | Medium | Low | Low | 2.7 | 2.5 |
| 19 | High | Medium | Low | 3.5 | 4.0 |

First of all, only 19 rules were generated because some combinations were absent from the learning data set. No vine plots were planted with a rootstock that confers either a Very Low or a Very High vine vigor level. After optimization, the fuzzy set parameters $C_2$ and $C_3$ of $VIG_S$ were identical (Table 1), so that there was no smooth transition between a Medium level of $VIG_S$ and a High level.

Consequents of rules 8 and 9 strongly decreased after optimization (-1.3 and -1.6 on a [1-4] scale) in contrast with the consequent of rule 2 that did not much change. For the rules corresponding to a Medium $VIG_S$, the rule conclusions systematically decreased after the optimization.

Table 1 (right) summarizes the results of optimization runs, comparing the average results of the initial and the median FIS over the learning and test samples. The median FIS significantly improved the accuracy over the test samples, with a relative gain of 19% for the RMSE and 22% for the $R^2$. The median FIS-

based model has a relatively good accuracy, so its behavior can be interpreted to study the interactions pointed out by the rules. Some examples are given below.

In rules 11, 13 and 19, environmental factors imparting a High vigor are not associated to a rootstock that confers a Low vigor level. Goulet and Morlat [17] already noticed that the practices in the vineyard are sometimes unsuitable because they have not been well adapted to environmental factors. For example, the authors indicate that in the vineyard of the Sarthe in Loire Valley (France), 72% of the vine plots have a too vigorous rootstock since the environmental factors already induce a very strong vigor.

The effect of the $VIG_C$ variable can also be discussed. When vine plots have no intercrop, i.e. no constraint on vine vigor, $VIG_C$=Low (rules 10, 12, 15, 17, 18 and 19), the estimated vigor is always higher than 2, unlike vine plots with an intercrop. The impact of a grass cover as intercrop on vine is well known in the literature due to competition for water and nitrogen. In [18], the authors indicated that intercrop reduces vine growth, i.e. the vigor, of the present year but also of the next years by decreasing grapevine nitrogen reserves. These already known relationships, interpreted by expertise, confirm the ability of the method to extract knowledge from a database.

Let us now consider plots intercropped with a crop that involves a High constraint ($VIG_C$=High). When the soil imparts a Medium or a Low vigor, the estimated vigor is coherent with the empirical knowledge: a Low vigor rootstock leads to a lower vigor; the more the soil imparts a Low vigor, the greater the difference between rootstocks. On the contrary, when the soil imparts a High vigor level, and for Low vigor rootstock, the system estimates a High vigor level (rule 7). At first sight, this is unexpected. It might be explained by the ability of some Low vigor rootstocks to adapt to soil humidity.

## 4 Conclusion

The use of a fuzzy formalism for inference systems increases the model complexity by introducing more parameters, by having to choose fuzzy operators and so on. One must be careful that this rise in complexity is accompanied by some benefits. The work presented in this paper tried to show the interest of using fuzzy inference systems that integrate expertise and data, for modeling a complex phenomenon in agronomy, namely vine vigor.

From the agronomical point of view, our procedure allows to study the combinations of features, therefore complementing the expertise, which is often related to the effect of one feature, independently from the other ones. This work lays down the foundations of a decision support tool aiming to adapt the agricultural practices to the environment in order to get a given vigor target. A next step consists in testing the method in other vineyards, including rule analysis and system behavior assessment.

From the methodological point of view, some work remains to be done to deal with the uncertainty of input and output measurements or assessments, for instance to define accuracy indices taking into account a fuzzy target instead of a crisp one.

## References

1. Guillaume, S., Charnomordic, B.: Fuzzy inference systems: an integrated modelling environment for collaboration between expert knowledge and data using fispro. Expert Systems with Applications 39, 8744–8755 (2012)
2. Thomopoulos, R., Destercke, S., Charnomordic, B., Iyan, J., Abécassis, J.: An iterative approach to build relevant ontology-aware data-driven models. Information Sciences 221, 452–472 (2013)
3. Sicard, M., Baudrit, C., Leclerc-Perlat, M.N., Wuillemin, P.H., Perrot, N.: Expert knowledge integration to model complex food processes. application on the camembert cheese ripening process. Expert Syst. Appl. 38, 11804–11812 (2011)
4. Guillaume, S., Charnomordic, B.: Learning interpretable fuzzy inference systems with fispro. International Journal of Information Sciences 181, 4409–4427 (2011)
5. Colin, F., Guillaume, S., Tisseyre, B.: Small catchment agricultural management using decision variables defined at catchment scale and a fuzzy rule-based system: a mediterranean vineyard case study. Water Resources Management 25, 2649–2668 (2011)
6. Rajaram, T., Das, A.: Modeling of interactions among sustainability components of an agro-ecosystem using local knowledge through cognitive mapping and fuzzy inference system. Expert Systems with Applications 37, 1734–1744 (2010)
7. Tremblay, N., Bouroubi, M., Panneton, B., Guillaume, S., Vigneault, P., Bélec, C.: Development and validation of fuzzy logic inference to determine optimum rates of n for corn on the basis of field and crop features. Precision Agriculture 11, 621–635 (2010)
8. Coulon-Leroy, C., Charnomordic, B., Rioux, D., Thiollet-Scholtus, M., Guillaume, S.: Prediction of vine vigor and precocity using data and knowledge-based fuzzy inference systems. Journal International des Sciences de la Vigne et du vin 46, 185–206 (2012)
9. Grelier, M., Guillaume, S., Tisseyre, B., Scholasch, T.: Precision viticulture data analysis using fuzzy inference systems. Journal International des Sciences de la Vigne et du vin 41, 19–31 (2007)
10. El Hajj, M., Bégué, A., Guillaume, S.: Integrating spot-5 time series, crop growth modeling and expert knowledge for monitoring agricultural practices - the case of sugarcane harvest on reunion island. Remote Sensing of Environment 113, 2052–2061 (2009)
11. Jones, H., Charnomordic, B., Dubois, D., Guillaume, S.: Practical inference with systems of gradual implicative rules. IEEE Transactions on Fuzzy Systems 17, 61–78 (2009)
12. de Oliveira, J.V.: Semantic constraints for membership functions optimization. IEEE Transactions on Systems, Man and Cybernetics. Part A 29, 128–138 (1999)
13. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Functions Algorithms. Plenum Press, New York (1981)
14. Guillaume, S., Charnomordic, B.: Parameter optimization of a fuzzy inference system using the fispro open source software. In: CFP12FUZ-USB, I.C.N: IEEE International Conference on Fuzzy Systems, pp. 402–409. IEEE, Brisbane (2012)

15. Glorennec, P.Y.: Algorithmes d'apprentissage pour systèmes d'inférence floue. Editions Hermès, Paris (1999)
16. Solis, F.J., Wets, R.J.: Minimization by random search techniques. Mathematics of Operation Research 6, 19–30 (1981)
17. Goulet, E., Morlat, R.: The use of surveys among wine growers in vineyards of the middle-loire valley (france), in relation to terroir studies. Land Use Policy 28, 770–782 (2010)
18. Celette, F., Findeling, A., Gary, C.: Competition for nitrogen in an unfertilized intercropping system: the case of an association of grapevine and grass cover ina mediterranean climate. European Journal of Agronomy 30, 41–51 (2009)

# On Fuzzy Polynomial Implications

Sebastia Massanet, Juan Vicente Riera, and Daniel Ruiz-Aguilera

University of the Balearic Islands
Crta. Valldemossa km 7.5, E-07122 Palma, Spain
{s.massanet,jvicente.riera,daniel.ruiz}@uib.es

**Abstract.** In this work, the class of fuzzy polynomial implications is introduced as those fuzzy implications whose expression is given by a polynomial of two variables. Some properties related to the values of the coefficients of the polynomial are studied in order to obtain a fuzzy implication. The polynomial implications with degree less or equal to 3 are fully characterized. Among the implications obtained in these results, there are some well-known implications such as the Reichenbach implication.

**Keywords:** Fuzzy implication, polynomial implication, (S,N)-implication, exchange principle.

## 1 Introduction

Fuzzy implications have become one of the most important operations in fuzzy logic. Their importance lies on the fact that they perform an analogous function to the classical implication in binary logic. Fuzzy implications generalize the classical ones in the sense that restricted to $\{0,1\}^2$ both coincide. Nowadays, these operations are modelled by means of monotonic functions $I : [0,1]^2 \to [0,1]$ satisfying the aforementioned border conditions. In the last years, a great number of researchers have devoted their efforts to the study of these logical connectives. Thus, we can highlight the survey [8] and the books [2] and [3], entirely devoted to fuzzy implications. This peak of interest in fuzzy implications is induced by the wide range of applications where these operations are useful. They play an essential role in approximate reasoning, fuzzy control, fuzzy mathematical morphology and other fields where these theories are applied.

All these applications trigger the need of having a large bunch of different classes of implications. In [11] the relevance of having many different classes of implications is pointed out. The main reason is that any "If-Then" rule can be modelled by a fuzzy implication and therefore, depending on the context and the proper behaviour of the rule, different implications can be suitable in each case. In addition, fuzzy implications are used to perform backward and forward inferences and so the choice of the implication can not be independent from the inference rule it is going to model.

In order to answer adequately to this necessity, several classes of fuzzy implications have been introduced. There exist two main strategies to obtain new classes.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 138–147, 2014.

The first one is based on the use of aggregation functions (t-norms, t-conorms, uninorms or aggregation functions in general) and other logical connectives, such as fuzzy negations. Some examples of this strategy are R and (S,N)-implications, QL and D-operations, among many others (see [3]). The second one is based on the use of univalued generators, obtaining the well-known Yager's implications or the $h$-implications. An exhaustive compilation of the different classes of fuzzy implications can be found in [10].

The implications obtained by means of these strategies can have very different expressions that will depend on the expressions of the aggregation functions or the generators used in their construction. However, the final expression of the fuzzy implication is important for its use in any application. It is well-known that some expressions of functions are tougher in order to compute their values and more propitious to spread possible errors caused by numerical approximations of the inputs. Consequently, operations with polynomial or rational expressions are more friendly than those which have a more complex expression from the computational point of view. Thus, in [1] and [7], all the rational Archimedean continuous t-norms are characterized. This family of t-norms is the well-known Hamacher class which contains the t-norms given by the following expression

$$T_\alpha(x,y) = \frac{xy}{\alpha + (1-\alpha)(x+y-xy)}, \quad x,y \in [0,1]$$

with $\alpha \geq 0$. Note that the only polynomial t-norm is the product t-norm $T_P(x,y) = xy$. Moreover, in [5], Fodor characterizes all the rational uninorms as those whose expression is given by

$$U_e(x,y) = \frac{(1-e)xy}{(1-e)xy + e(1-x)(1-y)}$$

if $(x,y) \in [0,1]^2 \backslash \{(0,1),(1,0)\}$ and, $U(1,0) = U(0,1) = 0$ or $U(0,1) = U(1,0) = 1$. In this case, there not exist any polynomial uninorm since they are never continuous.

So after recalling some definitions and results which will be used in this work, the main target is the introduction of fuzzy polynomial implications, those implications which have a polynomial of two variables as their expression. Some necessary conditions on the coefficients of a polynomial in order to be suitable to obtain a fuzzy implication are determined. After that, we will fully characterize all fuzzy polynomial implications of degree less or equal to 3 and we will study which additional properties they fulfil. From the derived results, the relationship of the obtained fuzzy polynomial implications with (S,N) and $f$-generated Yager's implications will be established. The paper ends with some conclusions and future work we want to develop.

## 2 Preliminaries

Let us recall some concepts and results that will be used throughout this paper. First, we give the definition of fuzzy negation.

**Definition 1 ([6, Definition 1.1]).** *A non-increasing function* $N : [0,1] \to [0,1]$ *is a* fuzzy negation*, if* $N(0) = 1$ *and* $N(1) = 0$*. A fuzzy negation* $N$ *is*

*(i)* strict*, if it is continuous and strictly decreasing.*
*(ii)* strong*, if it is an involution, i.e.,* $N(N(x)) = x$ *for all* $x \in [0,1]$*.*

Next, we recall the definition of fuzzy implication.

**Definition 2 ([6, Definition 1.15]).** *A binary operator* $I : [0,1]^2 \to [0,1]$ *is called a* fuzzy implication*, if it satisfies:*

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

From the definition, we can deduce that $I(0,x) = 1$ and $I(x,1) = 1$ for all $x \in [0,1]$ while the symmetrical values $I(x,0)$ and $I(1,x)$ are not determined. Some additional properties of fuzzy implications which will be used in this work are:

- The *left neutrality principle*,

$$I(1,y) = y, \quad y \in [0,1]. \tag{NP}$$

- The *exchange principle*,

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

- The *law of importation* with a t-norm $T$,

$$I(T(x,y),z) = I(x, I(y,z)), \quad x,y,z \in [0,1]. \tag{LI}$$

- The *ordering property*,

$$x \leq y \iff I(x,y) = 1, \quad x,y \in [0,1]. \tag{OP}$$

Finally, we recall the definitions of (S,N)-implications and Yager's f-generated implications.

**Definition 3 ([3, Definition 2.4.1]).** *A function* $I : [0,1]^2 \to [0,1]$ *is called an* (S,N)-implication *if there exist a t-conorm* $S$ *and a fuzzy negation* $N$ *such that*

$$I_{S,N}(x,y) = S(N(x),y), \quad x,y \in [0,1].$$

**Definition 4 ([3, Definition 3.1.1]).** *Let* $f : [0,1] \to [0,\infty]$ *be a continuous and strictly decreasing function with* $f(1) = 0$*. The function* $I_f : [0,1]^2 \to [0,1]$ *defined by*

$$I_f(x,y) = f^{-1}(x \cdot f(y)), \quad x,y \in [0,1],$$

*understanding* $0 \cdot \infty = 0$*, is called an* $f$-generated implication*. The function* $f$ *is an* $f$*-generator of the implication* $I_f$*.*

## 3 Polynomial Implications

In this section, we will introduce the concept of fuzzy polynomial implication and we will determine some necessary conditions on the coefficients of the polynomial in order to obtain a fuzzy implication from this expression.

*Remark 1.* Although in the introduction, the characterizations of rational Archimedean continuous t-norms and rational uninorms have been recalled (understanding a rational function as a quotient of two polynomials), in this work we will only focus on the fuzzy polynomial implications. This limitation is a direct consequence of the definition of a fuzzy implication. While uninorms and t-norms are associative functions and therefore, there exists a quite restrictive property in their definitions, fuzzy implications have a more flexible definition. This flexibility allows the existence of a great number of fuzzy polynomial implications and therefore, their study is worthy in itself.

**Definition 5.** *Let $n \in \mathbb{N}$. A binary operator $I : [0,1]^2 \to [0,1]$ is called a* fuzzy polynomial implication of degree $n$ *if it is a fuzzy implication and its expression is given by*

$$I(x,y) = \sum_{\substack{0 \leq i,j \leq n \\ i+j \leq n}} a_{ij} x^i y^j$$

*for all $x, y \in [0,1]$ where $a_{ij} \in \mathbb{R}$ and there exist some $0 \leq i, j \leq n$ with $i + j = n$ such that $a_{ij} \neq 0$.*

The following example shows the existence of fuzzy polynomial implications of any degree $n \in \mathbb{N}$ with $n \geq 2$.

*Example 1.* Let us consider the parametrized family of fuzzy negations given by $N_n(x) = 1 - x^n$ for all $x \in [0,1]$ and $n \in \mathbb{Z}^+$, and the probabilistic sum t-conorm, whose expression is $S_P(x,y) = x + y - xy$ for all $x, y \in [0,1]$. It is straightforward to check that the probabilistic sum belongs to the family of Hamacher t-conorms (the dual t-conorms of the Hamacher t-norms) and moreover, it is the unique polynomial t-conorm. Then, if we consider these two operators, we can construct the following parametrized family of (S,N)-implications

$$I_{S_P,N_{n-1}}(x,y) = S_P(N_{n-1}(x), y) = 1 - x^{n-1} + x^{n-1}y$$

for all $x, y \in [0,1]$ and $n \geq 2$. As it can be observed, they are polynomial implications of degree $n$. In addition, they satisfy (LI) with $T_P(x,y) = xy$ and therefore, they are also Yager's $f$-generated implications with $f(x) = \sqrt[n-1]{1-x}$ (see [9]).

A first property which can be derived form the definition is the continuity of these implications.

**Proposition 1.** *All fuzzy polynomial implications are continuous implications.*

*Remark 2.* It is worthy to note that some usual implications whose expression is piecewise polynomial will not be considered as polynomial functions. Thus, for instance, among many others, the following implications are not polynomial since they do not satisfy the requirements of the Definition 5:

$$I_{GD}(x,y) = \begin{cases} 1 \text{ if } x \leq y, \\ y \text{ if } x > y, \end{cases} \qquad I_{LK}(x,y) = \begin{cases} 1 & \text{if } x \leq y, \\ 1 - x + y \text{ if } x > y. \end{cases}$$

Note that both implications of the previous remark have a non-trivial one region. This fact is not a coincidence as the following result proves.

**Proposition 2.** *All fuzzy polynomial implications $I$ have a trivial one region, i.e., $I(x,y) = 1$ if, and only if, $x = 0$ or $y = 1$.*

This property is studied in detail in [4] where it is proved that this property is essential to generate strong equality indices. Furthermore, in particular, the polynomial implications never satisfy (OP). On the other hand, they can satisfy (EP) and in that case, they are (S,N)-implications.

**Proposition 3.** *Let $I(x,y) = \sum_{\substack{0 \leq i,j \leq n \\ i+j \leq n}} a_{ij} x^i y^j$ be a polynomial implication of degree $n$. If $I$ satisfies (EP), then $I$ is an (S,N)-implication generated by the strict fuzzy negation $N(x) = \sum_{i=0}^{n} a_{i0} x^i$ and the t-conorm $S(x,y) = \sum_{\substack{0 \leq i,j \leq n \\ i+j \leq n}} a_{ij} (N^{-1}(x))^i y^j$.*

However, the question of which polynomials can be fuzzy polynomial implications remains still unanswered. The problem relies on to characterize which coefficients $a_{ij} \in \mathbb{R}$ have to be chosen in order to generate a polynomial $p(x,y)$ satisfying the conditions of the Definition 2. We will partially answer this question in general for polynomials of degree $n$. First of all, the next result determines the necessary and sufficient conditions a polynomial must satisfy in order to be the expression of a fuzzy polynomial implication.

**Theorem 1.** *A polynomial $p(x,y) = \sum_{\substack{0 \leq i,j \leq n \\ i+j \leq n}} a_{ij} x^i y^j$ of degree $n$ is a fuzzy polynomial implication if, and only if, the following properties hold:*

*(i)* $p(0,y) = p(x,1) = 1$ *for all* $x, y \in [0,1]$.
*(ii)* $p(1,0) = 0$.
*(iii)* $\frac{\partial p(x,y)}{\partial x} \leq 0$ *for all* $x, y \in [0,1]$.
*(iv)* $\frac{\partial p(x,y)}{\partial y} \geq 0$ *for all* $x, y \in [0,1]$.
*(v)* $0 \leq p(1,y), p(x,0) \leq 1$.

The two first properties of the previous theorem provide some conditions on the coefficients $a_{ij}$ of the polynomial $p(x,y)$ in a direct way.

**Proposition 4.** *Let* $p(x,y) = \sum_{\substack{0\le i,j\le n\\ i+j\le n}} a_{ij}x^iy^j$ *be a polynomial of degree* $n$*. Then we have the following equivalences:*

*(i)* $p(0,y) = 1$ *if, and only if,* $a_{00} = 1$ *and* $a_{0j} = 0$ *for all* $0 < j \le n$.

*(ii)* $p(x,1) = 1$ *if, and only if,* $\sum_{j=0}^{n} a_{0j} = 1$ *and* $\sum_{j=0}^{n-i} a_{ij} = 0$ *for all* $0 < i \le n$.

*(iii)* $p(1,0) = 0$ *if, and only if,* $\sum_{i=0}^{n} a_{i0} = 0$.

Thus, the next result gives some necessary conditions on the coefficients of the fuzzy polynomial implications.

**Corollary 1.** *Let* $I(x,y) = \sum_{\substack{0\le i,j\le n\\ i+j\le n}} a_{ij}x^iy^j$ *be a polynomial implication of degree* $n$*. Then the following properties hold:*

*(i)* $a_{00} = 1$.

*(ii)* $a_{0j} = 0$ *for all* $0 < j \le n$.

*(iii)* $\sum_{j=0}^{n} a_{0j} = 1$ *and* $\sum_{j=0}^{n-i} a_{ij} = 0$ *for all* $0 < i \le n$.

*(iv)* $\sum_{i=1}^{n} a_{i0} = -1$.

However, the transfer of the properties (iii)-(v) of Theorem 1 to properties on the coefficients of the polynomial is harder for polynomials of degree $n$. Consequently, and with the aim of characterizing some polynomial implications, from now on we will restrict the study to polynomial implications of degree less or equal to 3.

### 3.1 Degree Less or Equal to One

First, we are going to study the existence of polynomial implications of degree less or equal to 2, i.e., fuzzy implications given by the following expression

$$\begin{aligned} &I(x,y) = a_{00}, && \text{with } a_{00} \in \mathbb{R},\\ &I(x,y) = a_{00} + a_{10}x + a_{01}y, && \text{with } a_{10} \neq 0 \text{ or } a_{01} \neq 0. \end{aligned}$$

It is easy to check that by Corollary 1 in the first case, it must hold that $a_{00} = 0$ and $a_{00} = 1$. Therefore, there not exist fuzzy polynomial implications of degree less or equal to 1. Let us recall again that there exist constant piecewise fuzzy implications. Two well-known examples of these implications are the least $I_{Lt}$ and the greatest $I_{Gt}$ fuzzy implications defined as follows

$$I_{Lt}(x,y) = \begin{cases} 1 \text{ if } x = 0 \text{ or } y = 1,\\ 0 \text{ otherwise,} \end{cases} \qquad I_{Gt}(x,y) = \begin{cases} 0 \text{ if } x = 1 \text{ and } y = 0,\\ 1 \text{ otherwise.} \end{cases}$$

Furthermore, there are also no fuzzy polynomial implications of degree 1. In this case, Corollary 1 states that the coefficients must satisfy $a_{00} = 1$, $a_{01} = 0$, $a_{10} = 0$ and also $a_{10} = -1$. Therefore, there is no feasible solution. However, there exist again fuzzy implications which are polynomial of degree less or equal to 1 piecewise. For instance, among many others, we have the fuzzy implications presented in Remark 2. To sum up, from the previous discussion, the following result is immediate.

**Proposition 5.** *There are no fuzzy polynomial implications of degree less or equal to 1.*

### 3.2 Degree 2

Now we deal with the characterization of all polynomial implications of degree 2, i.e., those whose expression is given by

$$I(x, y) = a_{00} + a_{10}x + a_{01}y + a_{11}xy + a_{20}x^2 + a_{02}y^2$$

with $a_{11}^2 + a_{20}^2 + a_{02}^2 \neq 0$. First of all, using Corollary 1, we obtain that there exists only a value for each coefficient, namely $a_{00} = 1$, $a_{10} = -1$, $a_{11} = 1$ and $a_{01} = a_{20} = a_{02} = 0$. Replacing these values into the expression of the polynomial, we get

$$p(x, y) = 1 - x + xy = I_{RC}(x, y),$$

i.e., we obtain the Reichenbach implication. Since it is a well-known fuzzy implication, it satisfies the remaining conditions of Theorem 1. Therefore, there exists only one fuzzy polynomial implication of degree 2.

**Proposition 6.** *There exists only one fuzzy polynomial implication of degree 2, the Reichenbach implication $I(x, y) = 1 - x + xy$.*

Note that the implication $I_{RC}$ is an (S,N)-implication obtained using the method introduced in Example 1 as $I_{S_P,N_1}$. It is well-known that this implication is also a Yager's $f$-generated implication with $f(x) = 1 - x$ and so, it satisfies (LI) with $T_P$.

### 3.3 Degree 3

Finally, in this section, we will analyse the fuzzy polynomial implications of degree 3. These implications have the following expression

$$I(x, y) = \sum_{\substack{0 \leq i,j \leq 3 \\ i+j \leq 3}} a_{ij} x^i y^j$$

where $a_{ij} \in \mathbb{R}$ and there exist $0 \leq i, j \leq 3$ with $i + j = 3$ such that $a_{ij} \neq 0$. Corollary 1 in this case provides some relations between the different coefficients.

**Corollary 2.** *Let* $I(x,y) = \sum_{\substack{0\leq i,j\leq 3\\ i+j\leq 3}} a_{ij}x^iy^j$ *be a fuzzy polynomial implication of degree 3. Then the following properties hold:*

- $a_{00} = 1$ *and* $a_{01} = a_{02} = a_{03} = a_{30} = 0$.
- $a_{12} = -a_{10} - a_{11}$.
- $a_{20} = -1 - a_{10}$.
- $a_{21} = 1 + a_{10}$.
- $a_{10} \neq -1$ *or* $a_{10} \neq -a_{11}$.

The previous result reduces the candidate polynomials to be the expression of a polynomial implication to

$$p(x,y) = 1 + a_{10}x + (-1 - a_{10})x^2 + a_{11}xy + (1 + a_{10})x^2y + (-a_{10} - a_{11})xy^2 \quad (1)$$

where $a_{10} \neq -1$ or $a_{10} \neq -a_{11}$. However, not all these polynomials satisfy the properties (iii)-(v) of Theorem 1 and therefore, not all are fuzzy implications. The next result fully characterizes all fuzzy polynomial implications of degree 3.

**Theorem 2.** *Let* $I : [0,1]^2 \to [0,1]$ *be a binary operator. Then* $I$ *is a fuzzy polynomial implication of degree 3 if, and only if,* $I$ *is given by*

$$I(x,y) = 1 + \alpha x + (-1 - \alpha)x^2 + \beta xy + (1 + \alpha)x^2y + (-\alpha - \beta)xy^2 \quad (2)$$

*with* $\alpha, \beta \in \mathbb{R}$, $\alpha \neq -1$, $\alpha \neq -\beta$, *and one of these cases hold:*

- $-2 \leq \alpha \leq -1$ *and* $-1 - \alpha \leq \beta \leq 2$.
- $-1 < \alpha < 0$ *and* $0 \leq \beta \leq -2\alpha$.
- $\alpha = \beta = 0$.

At this stage, let us study some properties of these implications in order to determine after that, the class of fuzzy implications which these operations belong to.

**Proposition 7.** *Let* $I$ *be a fuzzy polynomial implication of degree 3 given by Expression (2). Then the following statements are equivalent:*

- $I$ *satisfies (EP).*
- $I$ *satisfies (NP).*
- $\alpha = -\beta$ *with* $-2 \leq \alpha \leq 0$.

*In this case, the implication* $I$ *is given by*

$$I(x,y) = 1 + \alpha x + (-1 - \alpha)x^2 - \alpha xy + (1 + \alpha)x^2y. \quad (3)$$

Since the most usual fuzzy implications satisfy (NP), there exist fuzzy polynomial implications of degree 3 which are neither (S,N), R, QL nor D-implications. For example, the following fuzzy polynomial implications do not satisfy (NP)

$$I_1(x,y) = 1 - 2x + x^2 + xy - x^2y + xy^2, \quad I_2(x,y) = \frac{1}{2}(2 - x - x^2 + x^2y + xy^2).$$

On the other hand, using Proposition 3, the fuzzy polynomial implications of degree 3 satisfying (EP) are (S,N)-implications obtained from the unique polynomial t-conorm $S_P$.

**Theorem 3.** *Let $I : [0,1]^2 \to [0,1]$ be a binary operator, $S$ a t-conorm and $N$ a fuzzy negation. Then the following assertions are equivalent:*

*(i) $I$ is a fuzzy polynomial implication of degree 3 and an (S,N)-implication obtained from $S$ and $N$.*
*(ii) $S = S_P$ and $N(x) = 1 + \alpha x + (-1-\alpha)x^2$ with $-2 \leq \alpha \leq 0$.*

Finally, and using the recent characterization of Yager's $f$-generated implications in [9], the next result determines which fuzzy polynomial implications of degree 3 are Yager's implications.

**Theorem 4.** *Let $I : [0,1]^2 \to [0,1]$ be a binary operator. Then the following assertions are equivalent:*

*(i) $I$ is a fuzzy polynomial implication of degree 3 and a Yager's $f$-generated implication.*
*(ii) $I$ is a fuzzy polynomial implication of degree 3 satisfying (LI) with $T_P(x,y) = xy$.*
*(iii) $I(x,y) = 1 - x^2 + x^2 y$, the $f$-generated implication with $f(x) = \sqrt{1-x}$.*

As one might expect, the obtained implication belongs to the family constructed in Example 1 taking $I_{S_P,N_2}$.

## 4 Conclusions and Future Work

In this paper, we have started the study of fuzzy implications according to their final expression instead of the usual study on the construction methods of these operators using aggregation functions or generators. As a first step, we have studied the fuzzy polynomial implications, presenting some general results for polynomial implications of any degree and characterizing all fuzzy polynomial implications of degree less or equal to 3. The family of polynomial implications has a non-empty intersection with (S,N)-implications and $f$-generated implications, although there are also implications of this family which do not belong to any of the most usual families of implications. From the obtained results, some questions remain unanswered and must be tackled as future work. First,

*Problem 1.* Characterize all fuzzy polynomial implications of any degree.

For this purpose, it will be a vital requirement to determine which conditions on the coefficients of the polynomial imply the properties 3-5 of Theorem 1. Finally, from the results obtained in Proposition 6 and Theorem 3, we can conclude that all fuzzy polynomial implications of degree 2 or 3 which are also (S,N)-implications satisfy $S = S_P$. From this previous discussion, the next question emerges:

*Problem 2.* Is there any fuzzy polynomial implication which is also an (S,N)-implication obtained from a t-conorm $S \neq S_P$?

Finally, it would be interesting to check the advantages of using polynomial fuzzy implications instead of other implications in a concrete application in terms of computational cost saving and the reduction of the spreading of possible errors caused by numerical approximations of the inputs.

**Acknowledgments.** This paper has been partially supported by the Spanish Grants MTM2009-10320, MTM2009-10962 and TIN2013-42795-P with FEDER support.

## References

1. Alsina, C., Frank, M., Schweizer, B.: Associative functions. Triangular norms and copulas. World Scientific Publishing Company, Singapore (2006)
2. Baczyński, M., Beliakov, G., Bustince, H., Pradera, A.: Advances in Fuzzy Implication Functions. STUDFUZZ, vol. 300. Springer, Heidelberg (2013)
3. Baczyński, M., Jayaram, B.: Fuzzy Implications. STUDFUZZ, vol. 231. Springer, Heidelberg (2008)
4. Bustince, H., Fernandez, J., Sanz, J., Baczyński, M., Mesiar, R.: Construction of strong equality index from implication operators. Fuzzy Sets and Systems 211, 15–33 (2013)
5. Fodor, J.: On rational uninorms. In: Proceedings of the First Slovakian–Hungarian Joint Symposium on Applied Machine Intelligence, Herlany, Slovakia, pp. 139–147 (February 2003)
6. Fodor, J.C., Roubens, M.: Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Academic Publishers, Dordrecht (1994)
7. Hamacher, H.: Ein axiomatischer Beiitrag zur normativen Entscheidungstheorie. In: Über logische Aggregationen nicht binär explizierter Entscheidungskriterien. Verlag, Frankfurt (1978)
8. Mas, M., Monserrat, M., Torrens, J., Trillas, E.: A survey on fuzzy implication functions. IEEE Transactions on Fuzzy Systems 15(6), 1107–1121 (2007)
9. Massanet, S., Torrens, J.: On the characterization of Yager's implications. Information Sciences 201, 1–18 (2012)
10. Massanet, S., Torrens, J.: An overview of construction methods of fuzzy implications. In: Baczynski, M., Beliakov, G., Bustince, H., Pradera, A. (eds.) Adv. in Fuzzy Implication Functions. STUDFUZZ, vol. 300, pp. 1–30. Springer, Heidelberg (2013)
11. Trillas, E., Mas, M., Monserrat, M., Torrens, J.: On the representation of fuzzy rules. Int. J. Approx. Reasoning 48(2), 583–597 (2008)

# Implications Satisfying the Law of Importation with a Given Uninorm

Sebastia Massanet and Joan Torrens

University of the Balearic Islands
Crta. Valldemossa km 7.5, E-07122 Palma, Spain
{s.massanet,jts224}@uib.es

**Abstract.** In this paper a characterization of all fuzzy implications with continuous $e$-natural negation that satisfy the law of importation with a given uninorm $U$ is provided. The cases when the considered uninorm $U$ is representable or a uninorm in $\mathcal{U}_{\min}$ are studied separately and detailed descriptions of those implications with continuous natural negation with respect to $e$ that satisfy the law of importation with a uninorm in these classes are done. In the process some important examples are included.

**Keywords:** Fuzzy implication, law of importation, uninorm, fuzzy negation.

## 1 Introduction

Fuzzy implication functions are the generalization of binary implications in classical logic to the framework of fuzzy logic. Thus, they are used in fuzzy control and approximate reasoning to perform fuzzy conditionals [15, 20, 26] and also to perform forward and backward inferences in any fuzzy rules based system through the inference rules of modus ponens and modus tollens [17, 26, 30].

Moreover, fuzzy implication functions have proved to be useful in many other fields like fuzzy relational equations [26], fuzzy DI-subsethood measures and image processing [7, 8], fuzzy morphological operators [13, 14, 21] and data mining [37], among others. In each one of these fields, there are some additional properties that the fuzzy implication functions to be used should have to ensure good results in the mentioned applications.

The analysis of such additional properties of fuzzy implication functions usually reduces to the solution of specific functional equations. Some of the most studied properties are:

a) *The modus ponens*, because it becomes crucial in the inference process through the compositional rule of inference (CRI). Some works on this property are [23, 34–36].
b) *The distributivity properties* over conjunctions and disjunctions. In this case, these distributivities allow to avoid the combinatorial rule explosion in fuzzy systems (see [10]). They have been extensively studied again by many authors, see [1, 2, 4, 6, 31–33].

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 148–157, 2014.

c) *The law of importation.* This property is extremely related to the exchange principle (see [27]) and it has proved to be useful in simplifying the process of applying the CRI in many cases, see [3] and [16]. It can be written as

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

where $T$ is a t-norm (or a more general conjunction) and $I$ is a fuzzy implication function. The law of importation has been studied in [3, 16, 24, 25, 27]. Moreover, in this last article the law of importation has also been used in new characterizations of some classes of implications like $(S,N)$-implications and $R$-implications. Finally, it is a crucial property to characterize Yager's implications (see [28]).

Although all these works devoted to the law of importation, there are still some open problems involving this property. In particular, given any t-norm $T$ (conjunctive uninorm $U$), it is an open problem to find all fuzzy implications $I$ such that they satisfy the law of importation with respect to this fixed t-norm $T$ (conjunctive uninorm $U$). Recently, the authors have studied this problem, for implications with continuous natural negation, in the cases of the minimum t-norm and any continuous Archimedean t-norm (see [29]).

In this paper we want to deal with this problem but for the case of a conjunctive uninorm $U$ lying in the classes of representable uninorms and uninorms in $\mathcal{U}_{\min}$. We will give some partial solutions (in the sense that we will find all solutions involving fuzzy implications with an additional property). Specifically, we will characterize all fuzzy implication functions with continuous natural negation with respect to $e$ that satisfy the law of importation with any conjunctive uninorm $U$ in the mentioned classes. Along the process, some illustrative examples as well as particular cases when the fixed conjunctive uninorm $U$ is an idempotent uninorm in $\mathcal{U}_{\min}$ are presented separately.

## 2 Preliminaries

We will suppose the reader to be familiar with the theory of t-norms and t-conorms (all necessary results and notations can be found in [22]) and uninorms (see [12] and Chapter 5 in [3]). To make this work self-contained, we recall here some of the concepts and results used in the rest of the paper.

We will only focus on conjunctive uninorms in $\mathcal{U}_{\min}$ and representable uninorms.

**Theorem 1 ([12]).** *Let $U$ be a conjunctive uninorm with neutral element $e \in ]0,1[$ having functions $x \mapsto U(x,1)$ and $x \mapsto U(x,0)$ $(x \in [0,1])$ continuous except (perhaps) at the point $x = e$. Then $U$ is given by*

$$U(x,y) = \begin{cases} eT\left(\frac{x}{e}, \frac{y}{e}\right) & \text{if } (x,y) \in [0,e]^2, \\ e + (1-e)S\left(\frac{x-e}{1-e}, \frac{y-e}{1-e}\right) & \text{if } (x,y) \in [e,1]^2, \\ \min(x,y) & \text{otherwise.} \end{cases} \tag{1}$$

*where $T$ is a t-norm and $S$ is a t-conorm. In this case we will denote the uninorm by $U \equiv \langle T, e, S\rangle_{\min}$.*

The class of all uninorms with expression (1) will be denoted by $\mathcal{U}_{\min}$. Next, we give the definition of a conjunctive representable uninorm.

**Definition 1 ([12]).** *A conjunctive uninorm $U$ with neutral element $e \in (0,1)$ is representable if there exists a continuous and strictly increasing function $h : [0,1] \to [-\infty, +\infty]$ (called additive generator of $U$), with $h(0) = -\infty$, $h(e) = 0$ and $h(1) = +\infty$ such that $U$ is given by*

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

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

Now, we give some definitions and results concerning fuzzy negations.

**Definition 2 ([11, Definition 1.1]).** *A decreasing function $N : [0,1] \to [0,1]$ is called a fuzzy negation, if $N(0) = 1$, $N(1) = 0$. A fuzzy negation $N$ is called*

*(i) strict, if it is strictly decreasing and continuous,*
*(ii) strong, if it is an involution, i.e., $N(N(x)) = x$ for all $x \in [0,1]$.*

Next lemma plays an important role in the results presented in this paper. Essentially, given a fuzzy negation, it defines a new fuzzy negation which in some sense can perform the role of the inverse of the original negation.

**Lemma 1 ([3, Lemma 1.4.10]).** *If $N$ is a continuous fuzzy negation, then the function $\mathfrak{R}_N : [0,1] \to [0,1]$ defined by*

$$\mathfrak{R}_N(x) = \begin{cases} N^{(-1)}(x) & \text{if } x \in (0,1], \\ 1 & \text{if } x = 0, \end{cases}$$

*where $N^{(-1)}$ stands for the pseudo-inverse of $N$ given by $N^{(-1)}(x) = \sup\{z \in [0,1] \mid N(z) > x\}$ for all $x \in [0,1]$, is a strictly decreasing fuzzy negation. Moreover,*

*(i) $\mathfrak{R}_N^{(-1)} = N$,*
*(ii) $N \circ \mathfrak{R}_N = id_{[0,1]}$,*
*(iii) $\mathfrak{R}_N \circ N|_{Ran(\mathfrak{R}_N)} = id|_{Ran(\mathfrak{R}_N)}$,*

*where $Ran(\mathfrak{R}_N)$ stands for the range of function $\mathfrak{R}_N$.*

Now, we recall the definition of fuzzy implications.

**Definition 3 ([11, Definition 1.15]).** *A binary operator $I : [0,1]^2 \to [0,1]$ is said to be a fuzzy implication if it satisfies:*

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

Note that, from the definition, it follows that $I(0,x) = 1$ and $I(x,1) = 1$ for all $x \in [0,1]$ whereas the symmetrical values $I(x,0)$ and $I(1,x)$ are not derived from the definition. Fuzzy implications can satisfy additional properties coming from tautologies in crisp logic. In this paper, we are going to deal with the law of importation, already presented in the introduction.

The natural negation with respect to $e$ of a fuzzy implication will be also useful in our study.

**Definition 4 ([3, Definition 5.2.1]).** *Let $I$ be a fuzzy implication. If $I(1,e) = 0$ for some $e \in [0,1)$, then the function $N_I^e : [0,1] \to [0,1]$ given by $N_I^e(x) = I(x,e)$ for all $x \in [0,1]$, is called the* natural negation *of $I$ with respect to $e$.*

*Remark 1.*

(i) If $I$ is a fuzzy implication, $N_I^0$ is always a fuzzy negation.

(ii) Given a binary function $F : [0,1]^2 \to [0,1]$, we will denote by $N_F^e(x) = F(x,e)$ for all $x \in [0,1]$ its $e$-horizontal section. In general, $N_F^e$ is not a fuzzy negation. In fact, it is trivial to check that $N_F^e$ is a fuzzy negation if, and only if, $F(x,e)$ is a non-increasing function satisfying $F(0,e) = 1$ and $F(1,e) = 0$.

## 3 On the Satisfaction of (LI) with a Given Conjunctive Uninorm $U$

In this section, the main goal is the characterization of all fuzzy implications with a continuous natural negation with respect to $e \in [0,1)$ which satisfy the Law of Importation (LI) with a fixed conjunctive uninorm $U$.

First of all, the first question which arises concerns if fixed a concrete conjunctive uninorm $U$, any fuzzy negation can be the natural negation with respect to some $e \in [0,1)$ of a fuzzy implication satisfying (LI) with $U$. The answer is negative since as the following result shows, there exists some dependence between the conjunctive uninorm $U$ and the natural negation of the fuzzy implication $I$ with respect to some $e \in [0,1)$. To characterize which fuzzy negations are compatible with a conjunctive uninorm $U$ in this sense, the following property will be considered:

$$\text{if } N(y) = N(y') \text{ for some } y, y' \in [0,1], \text{ then } N(U(x,y)) = N(U(x,y')) \quad \forall x \in [0,1]. \tag{2}$$

Note that any strict negation obviously satisfies the previous equation. However, there are many other negations, not necessarily strict, which satisfy this property as we will see in next sections. Note also that similar conditions on a negation $N$ as (2) were considered in [9].

On the other hand, the following proposition is straightforward to check.

**Proposition 1.** *Let $I : [0,1]^2 \to [0,1]$ be a binary function such that $N_I^e$ is a fuzzy negation for some $e \in [0,1)$. If $I$ satisfies (LI) with a conjunctive uninorm $U$, then $N_I^e$ and $U$ satisfy Property (2).*

Next result gives the expression of any binary function with $N_I^e$ a continuous fuzzy negation for some $e \in [0,1)$ satisfying (LI) with a conjunctive uninorm $U$. Note that the binary function only depends on the uninorm $U$ and its natural negation with respect to $e \in [0,1)$.

**Proposition 2.** *Let $I : [0,1]^2 \to [0,1]$ be a binary function with $N_I^e$ a continuous fuzzy negation for some $e \in [0,1)$ satisfying (LI) with a conjunctive uninorm $U$. Then*

$$I(x,y) = N_I^e(U(x, \mathfrak{R}_{N_I^e}(y))).$$

From now on, we will denote these implications generated from a conjunctive uninorm $U$ and a fuzzy negation $N$ by $I_{N,U}(x,y) = N(U(x, \mathfrak{R}_N(y)))$.

*Remark 2.* Instead of $\mathfrak{R}_{N_I^e}$, we can consider any function $N_1$ such that $N_1^{(-1)} = N_I^e$ and $N_I^e \circ N_1 = \mathrm{id}_{[0,1]}$. This is a straightforward consequence of the satisfaction of Property (2) in this case. Since $N_I^e(\mathfrak{R}_{N_I^e}(y)) = N_I^e(N_1(y))$, then using the aforementioned property, $N_I^e(U(x, \mathfrak{R}_{N_I^e}(y))) = N_I^e(U(x, N_1(y)))$ and therefore, $I_{N_I^e,U}$ can be computed using either $\mathfrak{R}_{N_I^e}$ or $N_1$.

Moreover, this class of implications satisfies (LI) with the same conjunctive uninorm $U$ from which they are generated.

**Proposition 3.** *Let $N$ be a continuous fuzzy negation and $U$ a conjunctive uninorm satisfying Property (2). Then $I_{N,U}$ satisfies (LI) with $U$.*

Now, we are in condition to fully characterize the binary functions $I$ with $N_I^e$ a continuous fuzzy negation for some $e \in [0,1)$ satisfying (LI) with a conjunctive uninorm $U$.

**Theorem 2.** *Let $I : [0,1]^2 \to [0,1]$ be a binary function with $N_I^e$ a continuous fuzzy negation for some $e \in [0,1)$ and $U$ a conjunctive uninorm. Then*

$$I \text{ satisfies (LI) with } U \Leftrightarrow N_I^e \text{ and } U \text{ satisfy Property (2) and } I = I_{N_I^e,U}.$$

Note that it remains to know when $N_I^e$ and $U$ satisfy Property (2). From now on, we will try given a concrete conjunctive uninorm $U$, to determine which fuzzy negations satisfy the property with $U$.

## 4 On the Satifaction of Property 2 for Some Uninorms

In the previous section, Proposition 1 shows that the conjunctive uninorm and the natural fuzzy negation with respect to some $e \in [0,1)$ of the fuzzy implication must satisfy Property (2). Consequently, given a fixed conjunctive uninorm $U$, in order to characterize all fuzzy implications with a continuous natural negation with respect to some $e \in [0,1)$ satisfying (LI) with $U$, we need to know which fuzzy negations are compatible with the conjunctive uninorm $U$. In this section, we will answer this question for some conjunctive uninorms presenting

for each one, which fuzzy negations can be considered and then finally, using the characterization given in Theorem 2, the expressions of these fuzzy implications can be retrieved easily.

First of all, we want to stress again that the goal of this paper is to characterize all fuzzy implications with a continuous natural negation with respect to some $e \in [0,1)$ satisfying (LI) with a concrete conjunctive uninorm $U$. Therefore, there are other implications satisfying (LI) with a conjunctive uninorm $U$ than those given in the results of this section. Of course, these implications must have non-continuous natural negations with respect to any $e \in [0,1)$ such that $I(1,e) = 0$. An example of a fuzzy implication having non-continuous natural negations with respect to any $e \in [0,1)$ such that $I(1,e) = 0$ is the least fuzzy implication.

**Proposition 4.** *Let $I_{Lt}$ be the greatest fuzzy implication given by*

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

*Then $I_{Lt}$ satisfies (LI) with any conjunctive uninorm $U$.*

Consequently, although $I_{Lt}$ satisfies (LI) with any conjunctive uninorm $U$, we will not obtain this implication in the next results since it has no continuous natural negation at any level $e \in [0,1)$.

### 4.1 Representable Uninorms

The first class of uninorms we are going to study is the class of representable uninorms. The following result shows that the fuzzy negation must be strict in order to satisfy Property (2) with a uninorm of this class.

**Proposition 5.** *If $U$ is a representable uninorm, then Property (2) holds if, and only if, $N$ is an strict fuzzy negation.*

At this point, we can characterize all fuzzy implications with a continuous natural negation with respect to some $e \in [0,1)$ satisfying (LI) with a representable uninorm $U$.

**Theorem 3.** *Let $I : [0,1]^2 \to [0,1]$ be a binary function with $N_I^e$ a continuous fuzzy negation for some $e \in [0,1)$ and let $h : [0,1] \to [-\infty, +\infty]$ an additive generator of a representable uninorm. Then the following statements are equivalent:*

*(i) $I$ satisfies (LI) with the conjunctive representable uninorm $U_h$.*
*(ii) $N_I^e$ is strict and $I$ is given by* $I(x,y) = N_I^e(U(x,(N_I^e)^{-1}(y))) =$

$$\begin{cases} N_I^e(h^{-1}(h(x) + h((N_I^e)^{-1}(y)))) & \text{if } (x,y) \notin \{(0,0),(1,1)\}, \\ 1 & \text{otherwise.} \end{cases}$$

Note that the implications obtained in the previous theorem are in fact $(U,N)$-implications derived from the negation $(N_I^e)^{-1}$ and the uninorm $(N_I^e)^{-1}$-dual of $U_h$. Moreover, in the case that $(N_I^e)^{-1}$ coincides with the negation associated to the representable uninorm $U_h$, the implication is also the $RU$-implication derived from $U_h$ (see [5]). Similar results with implications derived from t-norms were also obtained in [29], see also [18, 19].

### 4.2 Uninorms in $\mathcal{U}_{\min}$

The second class of uninorms which we want to study is the class of uninorms in $\mathcal{U}_{\min}$. We will restrict ourselves to the cases where the underlying t-norm and t-conorm are continuous Archimedean or idempotent. Therefore, we will consider four different cases.

► $\boldsymbol{U \equiv \langle T_M, e, S_M \rangle_{\min}}$**.** In a first step, we will consider the uninorm $U \equiv \langle T_M, e, S_M \rangle_{\min}$ where $T_M(x, y) = \min\{x, y\}$ and $S_M(x, y) = \max\{x, y\}$, which in addition to the class of $\mathcal{U}_{\min}$, it belongs also to the class of idempotent uninorms, those satisfying $U(x, x) = x$ for all $x \in [0, 1]$. In contrast with the representable uninorms, in this case we will have continuous non-strict negations satisfying Property (2) with these uninorms.

**Proposition 6.** *If $U \equiv \langle T_M, e, S_M \rangle_{\min}$, then Property (2) holds if, and only if, $N$ is a continuous fuzzy negation satisfying the following two properties:*

1. *There exists $\alpha \in (0, 1]$ such that $N(x) = 0$ for all $x \geq \alpha$.*
2. *If $N(x) = k$ for all $x \in [a, b]$ for some constant $k > 0$ then $a \geq e$ or $b \leq e$.*

Note that any strict fuzzy negation satisfies the previous properties. In addition, those non-strict continuous fuzzy negations whose non-zero constant regions do not cross $x = e$ satisfy also Property (2) with $U \equiv \langle T_M, e, S_M \rangle_{\min}$. From this result and using Theorem 2, the expressions of the fuzzy implications we are looking for can be easily obtained.

► $\boldsymbol{U \equiv \langle T, e, S_M \rangle_{\min}}$ **with a Continuous Archimedean t-norm $\boldsymbol{T}$.** Now we focus on the case when we consider an underlying continuous Archimedean t-norm in addition to the maximum t-conorm. In this case, many of the fuzzy negations which were compatible with the uninorm of the first case are not compatible now with the uninorm of the current case.

**Proposition 7.** *If $U \equiv \langle T, e, S_M \rangle_{\min}$ with $T$ a continuous Archimedean t-norm, then Property (2) holds if, and only if, $N$ is a continuous fuzzy negation satisfying that there exists some $\alpha \in [0, e]$ such that $N(x) = 1$ for all $x \leq \alpha$ and $N$ is strictly decreasing for all $x \in (\alpha, e)$.*

Of course, as we already know, strict fuzzy negations are compatible with these uninorms. Furthermore, when it is continuous but non-strict, the only constant region allowed in $[0, e]$ is a one region while in $[e, 1]$, the fuzzy negation can have any constant region. Again, using Theorem 2, we can obtain the expressions of the fuzzy implications with a continuous natural negation with respect to some $e \in [0, 1)$ satisfying (LI) with some of these uninorms.

► $\boldsymbol{U \equiv \langle T_M, e, S \rangle_{\min}}$ **with a Continuous Archimedean t-conorm $\boldsymbol{S}$.** In this third case, we analyse the case when we consider an underlying continuous

Archimedean t-conorm in addition to the minimum t-norm. In this case, in contrast to the second case, now the main restrictions on the constant regions are located in $[e, 1]$.

**Proposition 8.** *If $U \equiv \langle T_M, e, S\rangle_{\min}$ with $S$ a continuous Archimedean t-conorm, then Property (2) holds if, and only if, $N$ is a continuous fuzzy negation satisfying:*

1. *There exists $\alpha \in (0, 1]$ such that $N(x) = 0$ for all $x \geq \alpha$.*
2. *If $N(x) = k$ for all $x \in [a, b]$ for some constant $k > 0$ then $b \leq e$.*

As always strict fuzzy negations are compatible with these uninorms. Moreover, when the fuzzy negation is continuous but non-strict, the only constant region which could cross $x = e$ is the zero region while in $[0, e]$, the fuzzy negation can have any constant region. Finally, Theorem 2 can be applied to obtain the expressions of these implications.

► **$U \equiv \langle T, e, S\rangle_{\min}$ with a Continuous Archimedean t-norm $T$ and t-conorm $S$.** In this last case, we analyse the case when we consider an underlying continuous Archimedean t-norm $T$ and t-conorm $S$. This is the case where fewer fuzzy negations are compatible with the considered uninorm. In fact, only two special constant regions are allowed.

**Proposition 9.** *If $U \equiv \langle T, e, S\rangle_{\min}$ with $T$ a continuous Archimedean t-norm and $S$ a continuous Archimedean t-conorm, then Property (2) holds if, and only if, $N$ is a continuous fuzzy negation satisfying the following two properties:*

1. *There exist $\alpha \in [0, e]$ and $\beta \in [e, 1]$ with $\alpha < \beta$ such that $N(x) = 1$ for all $x \leq \alpha$ and $N(x) = 0$ for all $x \geq \beta$.*
2. *$N$ is strict for all $x \in (\alpha, \beta)$.*

Clearly, we retrieve strict fuzzy negations when $\alpha = 0$ and $\beta = 1$. As we can see, continuous non-strict fuzzy negations are also possible but only two constant regions (zero and one regions) are allowed. In order to get the expressions of these implications, Theorem 2 must be used.

## 5 Conclusions and Future Work

In this paper, we have characterized all fuzzy implications satisfying (LI) with a conjunctive uninorm $U$ when the natural negation of the implication with respect to some $e \in [0, 1)$ is continuous. Moreover, we have determined in particular the expression of these implications when the conjunctive uninorm $U$ belongs to the class of $\mathcal{U}_{\min}$ with some underlying continuous Archimedean or idempotent t-norm and t-conorm and to the class of representable uninorms.

As a future work, we want to study the remaining uninorms of the class of $\mathcal{U}_{\min}$ and some other classes such as idempotent uninorms. In addition, we want to establish the relation between the new class of implications introduced in this paper $I_{N,U}$ and $(U, N)$-implications.

**Acknowledgments.** This paper has been partially supported by the Spanish Grants MTM2009-10320 and TIN2013-42795-P with FEDER support.

## References

1. Baczyński, M.: On the distributivity of fuzzy implications over continuous and archimedean triangular conorms. Fuzzy Sets and Systems 161(10), 1406–1419 (2010)
2. Baczyński, M.: On the distributivity of fuzzy implications over representable uninorms. Fuzzy Sets and Systems 161(17), 2256–2275 (2010)
3. Baczyński, M., Jayaram, B.: Fuzzy Implications. STUDFUZZ, vol. 231. Springer, Heidelberg (2008)
4. Baczyński, M., Jayaram, B.: On the distributivity of fuzzy implications over nilpotent or strict triangular conorms. IEEE Transactions on Fuzzy Systems 17(3), 590–603 (2009)
5. Baczyński, M., Jayaram, B.: Intersections between some families of (U,N)- and RU-implications. Fuzzy Sets and Systems 167(1), 30–44 (2011); Special Issue: Aggregation techniques and applications Selected papers from AGOP 2009 dedicated to Jaume Casasnovas
6. Balasubramaniam, J., Rao, C.: On the distributivity of implication operators over T and S norms. IEEE Transactions on Fuzzy Systems 12, 194–198 (2004)
7. Bustince, H., Mohedano, V., Barrenechea, E., Pagola, M.: Definition and construction of fuzzy DI-subsethood measures. Information Sciences 176, 3190–3231 (2006)
8. Bustince, H., Pagola, M., Barrenechea, E.: Construction of fuzzy indices from fuzzy DI-subsethood measures: application to the global comparison of images. Information Sciences 177, 906–929 (2007)
9. Cignoli, R., Esteva, F., Godo, L., Montagna, F.: On a class of left-continuous t-norms. Fuzzy Sets Syst. 131(3), 283–296 (2002)
10. Combs, W., Andrews, J.: Combinatorial rule explosion eliminated by a fuzzy rule configuration. IEEE Transactions on Fuzzy Systems 6(1), 1–11 (1998)
11. Fodor, J.C., Roubens, M.: Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Academic Publishers, Dordrecht (1994)
12. Fodor, J.C., Yager, R.R., Rybalov, A.: Structure of uninorms. Int. J. of Uncertainty, Fuzziness and Knowledge-based Systems 5, 411–427 (1997)
13. González-Hidalgo, M., Massanet, S.: A fuzzy mathematical morphology based on discrete t-norms: fundamentals and applications to image processing. Soft Computing, 1–15 (December 2013)
14. González-Hidalgo, M., Massanet, S., Mir, A., Ruiz-Aguilera, D.: On the choice of the pair conjunction-implication into the fuzzy morphological edge detector. IEEE Transactions on Fuzzy Systems (2013) (submitted)
15. Gottwald, S.: A Treatise on Many-Valued Logic. Research Studies Press, Baldock (2001)
16. Jayaram, B.: On the law of importation $(x \wedge y) \to z \equiv (x \to (y \to z))$ in fuzzy logic. IEEE Transactions on Fuzzy Systems 16, 130–144 (2008)
17. Jayaram, B.: Rule reduction for efficient inferencing in similarity based reasoning. International Journal of Approximate Reasoning 48(1), 156–173 (2008)
18. Jenei, S.: New family of triangular norms via contrapositive symmetrization of residuated implications. Fuzzy Sets Syst. 110(2), 157–174 (2000)

19. Jenei, S.: How to construct left-continuous triangular norms–state of the art. Fuzzy Sets and Systems 143(1), 27–45 (2004)
20. Kerre, E., Huang, C., Ruan, D.: Fuzzy Set Theory and Approximate Reasoning. Wu Han University Press, Wu Chang (2004)
21. Kerre, E., Nachtegael, M.: Fuzzy techniques in image processing. STUDFUZZ, vol. 52. Springer, New York (2000)
22. Klement, E., Mesiar, R., Pap, E.: Triangular norms. Kluwer Academic Publishers, Dordrecht (2000)
23. Mas, M., Monserrat, M., Torrens, J.: Modus ponens and modus tollens in discrete implications. Int. J. Approx. Reasoning 49(2), 422–435 (2008)
24. Mas, M., Monserrat, M., Torrens, J.: The law of importation for discrete implications. Information Sciences 179, 4208–4218 (2009)
25. Mas, M., Monserrat, M., Torrens, J.: A characterization of (U,N), RU, QL and D-implications derived from uninorms satisfying the law of importation. Fuzzy Sets and Systems 161, 1369–1387 (2010)
26. Mas, M., Monserrat, M., Torrens, J., Trillas, E.: A survey on fuzzy implication functions. IEEE Transactions on Fuzzy Systems 15(6), 1107–1121 (2007)
27. Massanet, S., Torrens, J.: The law of importation versus the exchange principle on fuzzy implications. Fuzzy Sets and Systems 168(1), 47–69 (2011)
28. Massanet, S., Torrens, J.: On the characterization of Yager's implications. Information Sciences 201, 1–18 (2012)
29. Massanet, S., Torrens, J.: Implications satisfying the law of importation with a given t-norm. In: Bustince, H., Fernandez, J., Mesiar, R., Calvo, T. (eds.) Aggregation Functions in Theory and in Practise. AISC, vol. 228, pp. 423–432. Springer, Heidelberg (2013)
30. Ruan, D., Kerre, E.E. (eds.): Fuzzy IF-THEN Rules in Computational Intelligence: Theory and Applications. Kluwer Academic Publishers, Norwell (2000)
31. Ruiz-Aguilera, D., Torrens, J.: Distributivity and conditional distributivity of a uninorm and a continuous t-conorm. IEEE Transactions on Fuzzy Systems 14(2), 180–190 (2006)
32. Ruiz-Aguilera, D., Torrens, J.: Distributivity of residual implications over conjunctive and disjunctive uninorms. Fuzzy Sets and Systems 158, 23–37 (2007)
33. Trillas, E., Alsina, C.: On the law $[(p \wedge q) \to r] = [(p \to r) \vee (q \to r)]$ in fuzzy logic. IEEE Transactions on Fuzzy Systems 10(1), 84–88 (2002)
34. Trillas, E., Alsina, C., Pradera, A.: On MPT-implication functions for fuzzy logic. Rev. R. Acad. Cien. Serie A. Mat. 98, 259–271 (2004)
35. Trillas, E., Alsina, C., Renedo, E., Pradera, A.: On contra-symmetry and MPT condicionality in fuzzy logic. International Journal of Intelligent Systems 20, 313–326 (2005)
36. Trillas, E., del Campo, C., Cubillo, S.: When QM-operators are implication functions and conditional fuzzy relations. International Journal of Intelligent Systems 15, 647–655 (2000)
37. Yan, P., Chen, G.: Discovering a cover set of ARsi with hierarchy from quantitative databases. Information Sciences 173, 319–336 (2005)

# Laws of Contraposition and Law of Importation for Probabilistic Implications and Probabilistic S-implications

Michał Baczyński[1], Przemysław Grzegorzewski[2,3], and Wanda Niemyska[1]

[1] Institute of Mathematics, University of Silesia,
40-007 Katowice, ul. Bankowa 14, Poland
{michal.baczynski,wniemyska}@us.edu.pl
[2] Systems Research Institute, Polish Academy of Sciences
01-447 Warsaw, Newelska 6, Poland
[3] Faculty of Mathematics and Information Science, Warsaw University of Technology,
Koszykowa 75, 00-662 Warsaw, Poland
pgrzeg@ibspan.waw.pl

**Abstract.** Recently, Grzegorzewski [5–7] introduced two new families of fuzzy implication functions called probabilistic implications and probabilistic S-implications. They are based on conditional copulas and make a bridge between probability theory and fuzzy logic. In the same article [7] author gives a motivation to his idea and indicates some interesting connections between new families of implications and the dependence structure of the underlying environment. In this paper the laws of contraposition and the law of importation are studied for these families of fuzzy implications.

**Keywords:** Fuzzy implications, Probabilistic implications, Laws of contraposition, Law of importation, Functional equations.

## 1 Introduction

Fuzzy implications belong to the main logical operations in fuzzy logic. They generalize the classical implication, which takes values in the set $\{0, 1\}$, to the unit interval $[0, 1]$. These functions are not only essential for fuzzy logic systems and fuzzy control, but they also play a significant role in solving fuzzy relational equations, in fuzzy mathematical morphology and image processing, and in defining fuzzy subsethood. In the scientific literature one can find many families of fuzzy implications along with their properties and applications. For the overview of this class of functions see the monograph [2] and the very recent book [1].

Recently, Grzegorzewski [5–7] introduced two new families of fuzzy implications - probabilistic implications and probabilistic S-implications. He also examined basic properties of families of probabilistic implications and probabilistic

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 158–167, 2014.

S-implications, like the left neutrality property, the exchange principle, the identity principle and the ordering property.

In this paper we continue the investigations and we examine the laws of contraposition and the law of importation for these two families of fuzzy implications. It should be noted that contrapositive symmetry of a fuzzy implication with respect to a fuzzy negation plays an important role in approximate reasoning, deductive systems, decision support systems, formal methods of proof, etc. (cf. [3, 10]). Similarly, fuzzy implications satisfying law of importation have been found very useful in fuzzy relational inference mechanisms, since one can obtain an equivalent hierarchical scheme which significantly decreases the computational complexity of the system (see [9, 13]).

The paper is organized as follows. After this introduction, Section 2 contains some preliminaries, where we recall basic concepts and definitions which will be used in the paper. Sections 3 introduces laws of contraposition and the law of importation. In Sections 4 and 5 we discuss the laws of contraposition for probabilistic implications and probabilistic S-implications, respectively, while Section 6 and 7 are devoted to the law of importation for these two families of implications, respectively. Last section contains conclusion and postulates some open problems.

## 2 Preliminaries

In this section we introduce general definitions of fuzzy operators, which will be exploited henceforth: fuzzy implications, triangular norms and fuzzy negations. We also recall the main definitions of probabilistic implications and probabilistic S-implications.

As it was mentioned before, fuzzy implication functions are generalizations of the classical implication to fuzzy logic. According to the well-established fact that the fuzzy concepts have to generalize adequately the corresponding crisp concepts, the most accepted definition of fuzzy implication nowadays is the following one.

**Definition 2.1 ([2, Definition 1.1.1]).** *A function* $I\colon [0,1]^2 \to [0,1]$ *is called a fuzzy implication if it satisfies the following conditions:*

**(I1)** *$I$ is non-increasing with respect to the first variable,*
**(I2)** *$I$ is non-decreasing with respect to the second variable,*
**(I3)** $I(0,0) = I(1,1) = 1$ *and* $I(1,0) = 0$.

**Definition 2.2 (see [11]).** *An associative, commutative and non-decreasing operation* $T\colon [0,1]^2 \to [0,1]$ *is called a t-norm if it has the neutral element 1.*

**Definition 2.3 (see [4, p. 3], [11, Definition 11.3]).** *A non-increasing function* $N\colon [0,1] \to [0,1]$ *is called a fuzzy negation if* $N(0) = 1$, $N(1) = 0$. *A fuzzy negation $N$ is called*

*(i) strict if it is strictly decreasing and continuous;*
*(ii) strong if it is an involution, i.e.* $N(N(x)) = x$ *for all* $x \in [0,1]$.

**Definition 2.4 (see [2, Definition 1.4.15]).** *Let $I$ be a fuzzy implication. The function $N_I$ defined by $N_I(x) = I(x, 0)$, $x \in [0, 1]$, is called the natural negation of $I$.*

**Definition 2.5 (see [12]).** *A copula (specifically, a 2-copula) is a function $C\colon [0,1]^2 \to [0,1]$ which satisfies the following conditions:*

**(C1)** $C(x, 0) = C(0, y) = 0$, *for all* $x, y \in [0, 1]$,
**(C2)** $C(x, 1) = x$, *for all* $x \in [0, 1]$,
**(C3)** $C(1, y) = y$, *for all* $y \in [0, 1]$,
**(C4)** $C(x_2, y_2) - C(x_2, y_1) - C(x_1, y_2) + C(x_1, y_1) \geq 0$, *for all* $x_1, x_2, y_1, y_2 \in [0, 1]$ *such that* $x_1 \leq x_2, y_1 \leq y_2$.

*Example 2.6.* The following are some basic copulas investigated in the literature.

(a) $C(x, y) = M(x, y) = \min(x, y)$.
(b) $C(x, y) = W(x, y) = \max(x + y - 1, 0)$.
(c) $C(x, y) = \Pi(x, y) = x \cdot y$.
(d) Farlie-Gumbel-Morgenstern's family, FGM($\theta$), where $\theta \in [-1, 1]$:

$$C_\theta(x, y) = x \cdot y + \theta x \cdot y(1 - x)(1 - y).$$

(e) Ali-Mikhail-Haq's family, AMH($\theta$), where $\theta \in [-1, 1]$:

$$C_\theta(x, y) = \frac{x \cdot y}{1 - \theta(1 - x)(1 - y)}.$$

It can be shown that every copula is bounded by the so-called Fréchet-Hoeffding bounds, i.e., for any copula $C$ and for all $x, y \in [0, 1]$ it holds

$$W(x, y) \leq C(x, y) \leq M(x, y).$$

Finally in this section, we present the formulas for probabilistic implications and probabilistic S-implications, which are based on conditional copulas.

**Definition 2.7 ([7]).** *Let $C$ be a copula. A function $I_C\colon [0,1]^2 \to [0,1]$ given by*

$$I_C(x, y) = \begin{cases} 1, & x = 0 \\ \frac{C(x,y)}{x}, & x > 0 \end{cases}, \qquad x, y \in [0, 1],$$

*is called a probabilistic implication (based on copula $C$).*

**Definition 2.8 ([7]).** *Let $C$ be a copula. A function $\tilde{I}_C\colon [0,1]^2 \to [0,1]$ given by*

$$\tilde{I}_C(x, y) = C(x, y) - x + 1, \qquad x, y \in [0, 1],$$

*is called a probabilistic S-implication (based on copula $C$).*

*Remark 2.9.* Not every probabilistic implication is also a fuzzy implication - we need to add the condition **(I1)**, that $I_C$ is non-increasing with respect to the first variable (the rest of conditions from fuzzy implication's definition are satisfied for any probabilistic implication, see [7]). Among functions based on the above examples of copulas,

$$I_{\mathbf{M}}(x,y) = I_{\mathbf{GG}}(x,y) = \begin{cases} 1, & x \leq y \\ \frac{y}{x}, & x > y \end{cases},$$

i.e., $I_{\mathbf{M}} = I_{\mathbf{GG}}$ is the Goguen implication and

$$I_{\Pi}(x,y) = \begin{cases} 1, & x = 0 \\ y, & x > 0 \end{cases},$$

i.e., $I_{\Pi} = I_{\mathbf{D}}$ is the least (S,N)-implication (see [2]), but $I_W$ is not. Probabilistic implications based on copulas from family FGM($\theta$) are fuzzy implications only for $\theta \geq 0$, whereas probabilistic implications based on copulas from family AMH($\theta$) are fuzzy implications for all $\theta \in [-1,1]$. However, it is worth noting that any probabilistic S-implication is a fuzzy implication.

## 3 The Laws of Contraposition and the Law of Importation

One of the most important tautologies in the classical two-valued logic is the law of contraposition:

$$p \to q \equiv \neg q \to \neg p.$$

Since the classical negation satisfies the law of double negation ($\neg(\neg p) \equiv p$), the following laws are also tautologies in the classical logic:

$$\neg p \to q \equiv \neg q \to p,$$
$$p \to \neg q \equiv q \to \neg p.$$

A natural generalization of those classical tautologies to fuzzy logic is based on fuzzy negations and fuzzy implications and plays an important role in various applications of fuzzy implications.

**Definition 3.1.** *Let $I$ be a fuzzy implication and $N$ be a fuzzy negation.*

*(i) We say that $I$ satisfies the law of contraposition with respect to $N$, if*

$$I(x,y) = I(N(y),N(x)), \qquad x,y \in [0,1]. \tag{CP}$$

*(ii) We say that $I$ satisfies the law of left contraposition with respect to $N$, if*

$$I(N(x),y) = I(N(y),x), \qquad x,y \in [0,1]. \tag{L-CP}$$

*(iii) We say that $I$ satisfies the law of right contraposition with respect to $N$, if*

$$I(x, N(y)) = I(y, N(x)), \qquad x, y \in [0,1]. \tag{R-CP}$$

*If $I$ satisfies the (left, right) contrapositive symmetry with respect to $N$, then we also denote this by CP($N$) (respectively, by L-CP($N$), R-CP($N$)).*

We can easily observe that the three properties introduced in Definition 3.1 are equivalent when $N$ is a strong negation (see [2, Proposition 1.5.3]).

The equation

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

known as the law of importation, is another tautology in classical logic. The general form of the above equivalence is introduced in the following definition.

**Definition 3.2 ([9]).** *Let $I$ be a fuzzy implication and $T$ be a t-norm. $I$ is said to satisfy the law of importation with t-norm $T$, if*

$$I(x, I(y, z)) = I(T(x, y), z), \qquad x, y, z \in [0,1]. \tag{LI}$$

## 4 The Laws of Contraposition for Probabilistic Implications

We start our investigation with the following result.

**Lemma 4.1.** *Let $I_C$ be a probabilistic implication based on some copula $C$. The natural negation $N_{I_C}$ is the least fuzzy negation, i.e., for all $x \in [0,1]$ we get*

$$N_{I_C}(x) = N_{\mathbf{D1}}(x) = \begin{cases} 1, & x = 0 \\ 0, & x > 0 \end{cases}.$$

*Proof.* For any probabilistic implication $I_C$ and any $x \in [0,1]$ we get

$$N_{I_C}(x) = I_C(x, 0) = \begin{cases} 1, & x = 0 \\ \frac{C(x,0)}{x}, & x > 0 \end{cases} = \begin{cases} 1, & x = 0 \\ 0, & x > 0 \end{cases}.$$

□

Now we are able to investigate the laws of contraposition for probabilistic implications.

**Lemma 4.2.** *No probabilistic implication satisfies the laws of contraposition* (CP) *and* (L-CP) *with any negation $N$.*

*Proof.* Since $I_C$ satisfies the left neutrality principle (see [7, Lemma 9]) and $N_{I_C}$ is not a strong negation, from [2, Corollary 1.5.5] we get that $I_C$ does not satisfy (CP) with any fuzzy negation.

Similarly, since $N_{I_C}$ is not a continuous negation, from [2, Corollary 1.5.15] we get that $I_C$ does not satisfy (L-CP) with any fuzzy negation. □

**Lemma 4.3.** *Every probabilistic implication satisfies* (R-CP) *only with respect to the least fuzzy negation* $N_{D1}$.

*Proof.* One can easily check that any probabilistic implication satisfies (R-CP) with $N_{D1}$. Indeed,

$$I_C(x, N_{D1}(y)) = \begin{cases} 1, & x = 0 \\ \frac{C(x, N_{D1}(y))}{x}, & x > 0 \end{cases} = \begin{cases} 1, & x = 0 \\ \frac{C(x,1)}{x}, & x > 0 \text{ and } y = 0 \\ \frac{C(x,0)}{x}, & x > 0 \text{ and } y > 0 \end{cases}$$

$$= \begin{cases} 1, & x = 0 \\ 1, & x > 0 \text{ and } y = 0 \\ 0, & x > 0 \text{ and } y > 0 \end{cases} = \begin{cases} 1, & x = 0 \text{ or } (x > 0 \text{ and } y = 0) \\ 0, & x > 0 \text{ and } y > 0 \end{cases}$$

$$= I_C(y, N_{D1}(x)).$$

Now, since $I_C$ satisfies the left neutrality principle (see [7, Lemma 9]), from [2, Lemma 1.5.21] we conclude that $N_{D1}$ is the only such negation. □

## 5 The Laws of Contraposition for Probabilistic S-Implications

Similarly, as in the previous section, we start our investigation with the following result.

**Lemma 5.1.** *Let* $\tilde{I}_C$ *be a probabilistic S-implication based on some copula* $C$. *The natural negation* $N_{\tilde{I}_C}$ *is the classical strong negation, i.e., for all* $x \in [0, 1]$ *we get*

$$N_{\tilde{I}_C}(x) = N_{\mathbf{C}}(x) = 1 - x.$$

*Proof.* For any probabilistic S-implication $\tilde{I}_C$ and all $x \in [0, 1]$ we get

$$N_{\tilde{I}_C}(x) = \tilde{I}_C(x, 0) = C(x, 0) - x + 1 = 1 - x.$$

□

**Lemma 5.2.** *Let* $\tilde{I}_C$ *be a probabilistic S-implication. If* $\tilde{I}_C$ *satisfies the* (CP) *with respect to a fuzzy negation* $N$, *then* $N$ *is the strong classical negation* $N_{\mathbf{C}}$.

*Proof.* We know that any probabilistic S-implication satisfies the left neutrality principle (see [7, Lemma 21]), so in view of [2, Lemma 1.5.4 (v)] we conclude that $N = N_{\tilde{I}_C} = N_{\mathbf{C}}$. □

**Corollary 5.3.** *Probabilistic S-implication* $\tilde{I}_C$ *based on a copula* $C$ *satisfies* (CP) *(with respect to* $N_{\mathbf{C}}$*) if and only if* $C$ *satisfies the following equation*

$$C(x, y) - x + 1 = C(1 - y, 1 - x) + y, \tag{1}$$

*for all* $x, y \in [0, 1]$.

*Remark 5.4.* At this moment we are not able to solve the above equation, but it should be noted that some probabilistic S-implications $\tilde{I}_C$ satisfy the law of contraposition (CP) with respect to $N_{\mathbf{C}}$ (e.g. implications based on copulas $\Pi$, $M$, $W$ or copulas from the family FGM($\theta$)), and some do not (e.g. some implications based on copulas from the family AMH($\theta$)). In fact probabilistic S-implications were studied by Hliněná et al. [8] in connection with generator triples of fuzzy preference structures and their Proposition 7 is the other version of our Corollary 5.3.

*Example 5.5.* We show that some probabilistic implications and/or probabilistic S-implications based on copulas from the family AMH($\theta$) do not satisfy CP($N_{\mathbf{C}}$). Let us take any copula $C \in AMH(\theta)$ and assume that (1) holds, i.e.

$$\frac{xy}{1-\theta(1-x)(1-y)} - x + 1 = \frac{(1-x)(1-y)}{1-\theta xy} + y. \tag{2}$$

Let $\theta = 1$ and $x = y = \frac{1}{4}$. Then the left side of (2) equals $\frac{25}{28}$, while the right side of (2) is equal to $\frac{17}{20}$, which is a contradiction. In fact we can get the result, that condition (1) is satisfied if and only if $(\theta = 0 \vee x \in \{0,1\} \vee y \in \{0,1\} \vee x + y = 1)$.

**Lemma 5.6.** *Let $\tilde{I}_C$ be a probabilistic S-implication. If $\tilde{I}_C$ satisfies the* (L-CP) *or* (R-CP) *with respect to a fuzzy negation $N$, then $N$ is the strong classical negation $N_{\mathbf{C}}$.*

*Proof.* Since any probabilistic S-implication satisfies the left neutrality principle (see [7, Lemma 21]), in view of [2, Lemma 1.5.14] and [2, Lemma 1.5.21], respectively, we conclude that $N = N_{\tilde{I}_C} = N_{\mathbf{C}}$ in both cases. □

The last fact in this section, for which we omit the proof, indicates new big groups of probabilistic S-implications which appropriately satisfy (CP) and which do not satisfy (CP).

**Theorem 5.7.** *Let $C_1$, $C_2$ be some copulas. For any copula $C$ such that $C = \theta C_1 + (1-\theta)C_2$, where $\theta \in [0,1]$, the following equivalence holds:*

$$\begin{pmatrix} \tilde{I}_C \\ \textit{satisfies (CP)} \\ \textit{with respect to } N_{\mathbf{C}}. \end{pmatrix} \iff \begin{pmatrix} \textit{both } \tilde{I}_{C_1} \textit{ and } \tilde{I}_{C_2} \\ \textit{satisfy (CP)} \\ \textit{with respect to } N_{\mathbf{C}}. \end{pmatrix}$$

## 6 The Law of Importation (LI) for Probabilistic Implications

**Lemma 6.1.** *If a probabilistic implication $I_C$ and a t-norm $T$ satisfy the law of importation* (LI)*, then $T$ is positive, i.e., $\neg\exists_{x,y\neq 0} T(x,y) = 0$.*

*Proof.* For $x = 0$ or $y = 0$ it is easy to check, that (LI) is satisfied for any probabilistic implication $I_C$ and any t-norm $T$.

Now, let us assume, that $x, y \neq 0$ and $z = 0$. Then

$$I_C(x, I_C(y, 0)) = I_C(x, 0) = 0,$$

and

$$I_C(T(x,y), 0) = \begin{cases} 1, & T(x,y) = 0 \\ 0, & T(x,y) > 0 \end{cases}.$$

Therefore, if there exist $x, y \neq 0$ such that $T(x,y) = 0$, then $I_C(x, I_C(y,0)) \neq I_C(T(x,y), 0)$. □

Now we will study five examples of copulas and probabilistic implications based on these copulas, looking for the positive t-norms, for which those specific implications satisfy (LI). We can assume that $x, y, z \neq 0$. In this case (LI) is equivalent to

$$\frac{C(T(x,y), z)}{T(x,y)} = \frac{C(x, \frac{C(y,z)}{y})}{x}.$$

*Example 6.2.*

(i) Probabilistic implication $I_\Pi$ satisfies (LI) with any positive t-norm $T$.
(ii) Probabilistic implication $I_M$ satisfies (LI) only with the product t-norm $T_P$.
(iii) Probabilistic implication $I_W$ satisfies (LI) only with product t-norm $T_P$.
(iv) Probabilistic implications generated by copulas from the family FGM($\theta$) do not satisfy (LI) with any t-norm $T$ (except the case when $\theta = 0$, but then we deal with product copula $\Pi$ - see case (i)).
(v) Among probabilistic implications generated by copulas from the family AMH($\theta$) for $\theta \neq 0$, only implication for $\theta = 1$ satisfies (LI) with product t-norm $T_P$.

Let us notice another interesting fact.

*Remark 6.3.* Probabilistic implications generated by $\Pi$ and $M$ satisfy (LI) with the t-norm $T_P$, however no probabilistic implication generated by copulas $C = \theta\Pi + (1-\theta)M$ for $\in (0,1)$ satisfy (LI) (with $T_P$ or any other t-norm $T$).

## 7 The Law of Importation (LI) for Probabilistic S-Implications

**Lemma 7.1.** *If a probabilistic S-implication $\tilde{I}_C$ satisfies* (LI) *with a t-norm $T$, then $T$ must be of the form $T(x,y) = x - C(x, 1-y)$.*

*Proof.* Let $z = 0$. Then for any $x, y \in [0,1]$ we get

$$\begin{aligned} C(T(x,y),z) - T(x,y) + 1 = C(x, C(y,z) - y + 1) - x + 1 \\ \iff \quad 0 - T(x,y) + 1 = C(x, 0 - y + 1) - x + 1 \\ \iff \quad T(x,y) = x - C(x, 1-y). \end{aligned}$$

□

*Remark 7.2.* On the other endpoints of $x, y, z$'s domains ($x = 0$, $y = 0$, $x = 1$, $y = 1$ and $z = 1$) the equation (LI) is satisfied for any probabilistic S-implication $\tilde{I}_C$ and the t-norm $T(x,y) = x - C(x, 1-y)$.

What we have to do now is to check for every probabilistic S-implication $\tilde{I}_C$ if the function $T(x,y) = x - C(x, 1-y)$ satisfies (LI) on the rest of its domain, i.e. for $x, y, z \in (0,1)$, and if so defined $T$ is a t-norm.

*Example 7.3.* Table 1 includes the list of functions which are the only candidates for t-norms for which particular probabilistic S-implications may satisfy (LI).

**Table 1.** Copulas and candidates for t-norms (see Example 7.3)

| Copula | Candidate for (LI) |
|---|---|
| $C_{\Pi}$ | $T_P(x,y) = xy$ |
| $C_M$ | $T_{LK}(x,y) = \max(x+y-1, 0)$ |
| $C_W$ | $T_M(x,y) = \min(x,y)$ |
| $C \in FGM(\theta)$ | $T(x,y) = xy - \theta xy(1-x)(1-y)$ |
| $C \in AMH(\theta)$ | $T(x,y) = xy\dfrac{1-\theta(1-x)}{1-\theta(1-x)y}$ |

It can be proved that the pairs of functions $(\tilde{I}_{C_\Pi}, T_P), (\tilde{I}_{C_M}, T_{LK}), (\tilde{I}_{C_W}, T_M)$ satisfy (LI), whereas probabilistic S-implications generated by the copulas from families FGM($\theta$) or AMH($\theta$) do not satisfy (LI) with functions listed in Table 1.

Please note that in [8] the function $T$ from the above Lemma 7.1 is denoted as $C^{\text{flip2}}$. Moreover in [8, Proposition 6] it is shown that $C^{\text{flip2}}$ is commutative (necessary condition for $T$ to be a t-norm) if and only if $C = (\hat{C})^t$, where $\hat{C}$ is the survival quasicopula to $C$. In fact the condition $C = (\hat{C})^t$ is fulfilled if and only if $I(x,y) = C(x,y) + 1 - x$ (i.e., the probabilistic S-implication) possesses the (CP) property (with respect to $N_{\mathbf{C}}$).

## 8 Conclusion

In this paper we have examined particular properties of probabilistic implications and probabilistic S-implications, i.e. (CP), (L-CP), (R-CP) and (LI).

However, some questions are still open. For instance, is there any particular family of copulas which satisfy the equation (1) from Corollary 5.3? Or, does the formula $T(x, y) = x - C(x, 1 - y)$ from Lemma 7.1 express any special kind of relation between functions $T$ and $C$ (some "duality")? We may draw hypothesis that if any probabilistic implication satisfies (LI) with a t-norm $T$, then it also satisfies (LI) with the product t-norm $T_P$.

It should be stressed that recently introduced probabilistic implications and probabilistic S-implications are still not enough examined families of implications and there are many other interesting properties which should be considered in their case.

## References

1. Baczyński, M., Beliakov, G., Bustince, H., Pradera, A. (eds.): Advances in Fuzzy Implication Functions. STUDFUZZ, vol. 300. Springer, Heidelberg (2013)
2. Baczyński, M., Jayaram, B.: Fuzzy Implications. STUDFUZZ, vol. 231. Springer, Heidelberg (2008)
3. Fodor, J.C.: Contrapositive symmetry of fuzzy implications. Fuzzy Sets and Systems 69, 141–156 (1995)
4. Fodor, J.C., Roubens, M.: Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer, Dordrecht (1994)
5. Grzegorzewski, P.: Probabilistic implications. In: Proceedings of the 7th Conference on European Society for Fuzzy Logic and Technology (EUSFLAT-2011) and LFA-2011, pp. 254–258. Atlantis Press, Amsterdam (2011)
6. Grzegorzewski, P.: On the properties of probabilistic implications. In: Melo-Pinto, P., Couto, P., Serôdio, C., Fodor, J., De Baets, B. (eds.) Eurofuse 2011. AISC, vol. 107, pp. 67–78. Springer, Heidelberg (2011)
7. Grzegorzewski, P.: Probabilistic implications. Fuzzy Sets and Systems 226, 53–66 (2013)
8. Hliněná, D., Kalina, M., Král', P.: Fuzzy preference modelling and some consequences for fuzzy implications and quasi-copulas. In: Proceeding of the 8th Conference of the European Society for Fuzy Logic and Technology (EUSFLAT 2013), pp. 260–265. Atlantis Press, Amsterdam (2013)
9. Jayaram, B.: On the law of importation $(x \wedge y) \to z \equiv (x \to (y \to z))$ in fuzzy logic. IEEE Trans. Fuzzy Systems 16, 130–144 (2008)
10. Jenei, S.: A new approach for interpolation and extrapolation of compact fuzzy quantities. The one dimensional case. In: Klement, E.P., Stout, L.N. (eds.) Proc. 21st Linz Seminar on Fuzzy Set Theory, Linz, Austria, pp. 13–18 (2000)
11. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer, Dordrecht (2000)
12. Nelsen, R.B.: An Introduction to Copulas. Springer, New York (1999)
13. Štěpnička, M., Jayaram, B.: On the suitability of the Bandler-Kohout subproduct as an inference mechanism. IEEE Trans. Fuzzy Systems 18, 285–298 (2010)

# Sequential Clustering for Event Sequences and Its Impact on Next Process Step Prediction

Mai Le[1], Detlef Nauck[2], Bogdan Gabrys[1], and Trevor Martin[3]

[1] University of Bournemouth
[2] British Telecommunications
[3] University of Bristol

**Abstract.** Next step prediction is an important problem in process analytics and it can be used in process monitoring to preempt failure in business processes. We are using logfiles from a workflow system that record the sequential execution of business processes. Each process execution results in a timestamped event. The main issue of analysing such event sequences is that they can be very diverse. Models that can effectively handle diverse sequences without losing the sequential nature of the data are desired. We propose an approach which clusters event sequences. Each cluster consists of similar sequences and the challenge is to identify a similarity measure that can cope with the sequential nature of the data. After clustering we build individual predictive models for each group. This strategy addresses both the sequential and diverse characteristics of our data. We first employ K-means and extent it into a categorical-sequential clustering algorithm by combining it with sequential alignment. Finally, we treat each resulting cluster by building individual Markov models of different orders, expecting that the representative characteristics of each cluster are captured.

## 1 Introduction

In order to achieve operational excellence, companies must run efficient and effective processes [1], [2]. They must also be able to predict if processes will complete successfully or run into exceptions in order to intervene at the right time, preempt problems and maintain customer service.

It is a real challenge to build such models for business process data due to many reasons. Let us point out two main reasons which we think most dominant. First, business process data is sequential in nature. A business process instance ($S_j$) is a composition of discrete events (or tasks) in a time-ordered sequence, $S_j = \left\{s_1^{(j)}, s_2^{(j)}, \ldots, s_{n_j}^{(j)}\right\}$, $s_k$ takes values from a finite set of event types $E = \{e_1, \ldots, e_L\}$. Each of these events has its own attributes. For simplicity, we assume that a process does not contain any overlapping events, that means there are no parallel structures [1]. Second, business process data can be very diverse because in practice, processes are typically designed based on knowledge about how a certain objective can be achieved efficiently. When process execution is not enforced by automatic workflow systems, people do not always follow the

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 168–178, 2014.

design. In large companies, many departments can be involved in the execution or maintenance of a process and processes can be implemented across a large number of IT systems. In these environments it can easily happen that over time the company loses track of the original process design and process evolves in an uncontrolled fashion. Consequently, there are many prototypes (different execution sequences) for one process [1].

Even though there is a rich source of mathematical models in data mining, not any sequential approaches seem to be effective in solving this particular problem. The solution is to 'divide' the process into smaller groups of tasks/steps and at each group, build a model accordingly [3], [4]. The shortcoming of these local models is that the coherence and the interaction between events from different event logs are lost. We propose another strategy addressing the complexity and diversity of process data which partitioning process data into groups of sequences of similar characteristics. Mathematical models are chosen according to the properties of the sequences in the resulting groups. The strength of this method is that it keeps the sequential form of the process, discriminates and adequately deals with different representative prototypes.

Existing sequential clustering approaches are found in the works of [5], [6] etc. The principle of these approaches is to build a distance measure matrix by first, modelling the data sequences one by one then comparing the likelihood of a sequence fitting to a chosen model. Any probabilistic model can be used here to describe the data e.g. linear autoregressive, graph-based models etc. HMM based sequential clustering is the most common and has shown its great performance in certain fields where data consists of continuous and/or long sequences. To fit data sequences to descriptive models, the data is assumed to have some properties or prior probabilistic distribution. It might be a better idea comparing the sequences directly based on the events and the order in which they occurred than to build a HMM for each sequence in our data, in particular for short length sequences. Hence, we use local sequence alignment (SA) to match all sequences pairwise and the outcome of the matchings are used as a similarity measure function. The proposed sequential clustering algorithm provides subgroups of sequences which are similar in the way events have occurred. Each resulting cluster is then treated by a couple of hybrid Markov models which are studed in our previous work.

The rest of the paper is organised as follows: Section 2 presents the sequence alignment technique. It is followed by Section 3 which introduces clustering and the proposed sequential clustering approach. Experiments and experimental results discussion take place in Section 4. Finally, Section 5 will conclude and draw a future research plan based on hints obtained from the former section.

## 2 Sequence Alignment - Similarity Measure Function

Algorithms used in sequence alignment are mainly categorised into global alignment and local alignment. Global alignment provides a global optimisation solution, which spans the entire length of all query sequences. One such algorithm

was introduced by Needleman and Wunchs [7]. In contrast, local alignment aims to find the most similar segments from two query sequences [8], [9]. In this work we use local alignment. There are two basic matrices associated with local sequence alignment algorithms: substitution matrix and score matrix. The role of the substitution matrix is to generate the degree of matching any two events from the set of event types, or in other words matching subsequences of length 1. This degree which is irrespective of the position of the events then contributes to the matching score in the score matrix that considers the complete sequences, i.e. all events in the order they occur. We now introduce these two matrices.

**Substitution Matrix:** In biology a substitution matrix describes the rate at which one amino acid in a sequence transforms to another amino acid over time. Regarding business process data no mutation occurs. Therefore, we do not need the complex version of the substitution matrix and we use the identity matrix.

**Score Matrix:** This matrix's elements are similarity degrees of events from the two given sequences considering the positions.

$$h_{i0} = h_{0j} = h_{00} = 0, \tag{1}$$

These $h_{i0}$, $h_{0j}$ and $h_{00}$ values are the initial values for the recursive formula that is used to compute $h_{ij}$.

$$\begin{aligned} h_{ij} = \max \{ & h_{i-1,j} - \delta, h_{i-1,j-1} + s(x_i, y_j), \\ & h_{i,j-1} - \delta, 0 \}, \end{aligned} \tag{2}$$

where $s(x_i, y_j)$ is the element of the substitution matrix and $x_i$, $y_j$ are events at positions $i$ and $j$. $\delta$ is the penalty for deletion or insertion. The $i^{th}$ event in a sequence can be aligned to the $j^{th}$ event in another sequence, or can be aligned to nothing (deletion). The optimal pair of aligned segments in this case is identified by the highest score in the matrix. The segment is then found by tracking back from that optimal highest score diagonally up toward the left corner until 0 is reached.

## 3 Data Clustering

Clustering is one of the main constituent elements in data mining. It is known as an unsupervised learning family. The aim of data clustering is to get the data distributed into a finite number of clusters, typically based on the distance between data points. Hence, a distance measure function is required and is vitally important. Clustering aims to group data in a way that each object in one cluster is similar to other objects in the same cluster more than to other objects in other clusters. Clustering approaches mainly are one of the following types:

- hierarchical clustering: agglomerative clustering [10], [11],
- probabilistic clustering: EM algorithm [12],
- partitioning clustering: K means clustering, K modes, K prototypes [10],

- fuzzy clustering [13],
- grid based clustering algorithm [14] (no distance required, only population of the grids is counted),
- graph-based algorithm Click [15].

As sometimes sequential features describe data best, one option is to consider each sequence as a multivariate feature vector and use vector composition based clustering to cluster the given sequences [16]. However, decomposition based approaches require sequences of the same length. Overcoming such issue, there are a large number of HMM-based sequential clustering algorithms and extensions [5], [6], [17], [18] etc. In these publications, the authors model individual data sequences by probabilistic models then use likelihood to build a distance matrix. Traditional clustering techniques are applied to partitioning the data using the obtained distance matrix.

$$d_{ij} = \frac{l(s_i, \lambda_j) + l(s_j, \lambda_i)}{2}, \tag{3}$$

where $d_{ij}$ are the distance between sequence $i$ and sequence $j$, $l(s_i, \lambda_j)$ is the likelihood of a sequence $i$ belonging to a model $\lambda(j)$.

**K Means Clustering.** K means clustering in data mining is itself an NP hard problem [19]. However, heuristic K means algorithms exist and provide locally optimal solutions. Some common K means algorithms are Lloyd algorithm [19], Bradley, Fayyad and Reina algorithm [10]. K means is widely used because of its simplicity and competence. Researchers have tried to improve the original approach. One of the alternative algorithms is to modify the original K means to profit from extra information about some specific data points should or should not be assigned to certain clusters [20] (constraints of belonging). Another alternative algorithm is to make the algorithm faster by using triangle inequality to avoid unnecessary computation [21].

$K$ is usually chosen in the first place or estimated by trial but there are also a number of studies on how to select a reasonable value for $K$ [22]. Given the number of clusters $K$, first, the corresponding centers are initialised. These centers can be data points randomly taken from the available dataset. Second, each data point is assigned to the closest cluster based on the distance between the data point and $K$ centers. Once all the data points are assigned, a new center for each cluster is determined. For numerical data, such centers are mean values of the elements in the corresponding clusters. The procedure of assigning data points to clusters and recomputing centers is repeated until it converges.

**K Means Variant for Event Sequences.** The sequence matching degree presented earlier is used as similarity measure in our proposed K means clustering which will be called SA based clustering from now on through out this paper. Because we directly use ordered events of data sequences to compare sequences, there is no mean value for each cluster. We resort to choose the sequence in each cluster whose total distance (similarity) to other members of the cluster is smallest (largest) as the new center.

## 4 Evaluation

We carried out a number of experiments based on records from two real processes ($DS1-2$) from a multi-national telecommunications company. $DS1$ is a legacy process, it consists of 2367 process instances of different lengths, varying from 1 to 78, 285 unique tasks (events) and about 20 process attributes. $DS2$ is also a real process with 11839 entries, 576 process instances with different lengths, and also has hundreds of unique tasks. The lengths of process instances in $DS2$ vary considerably. We would like to illustrate our process data and its complex property as mentioned in the introduction before getting into the main experiments. The diagram in Figure 1 shows the complexity of the subset of $DS1$. It is basically impossible to visually analyse or understand the process from this figure.

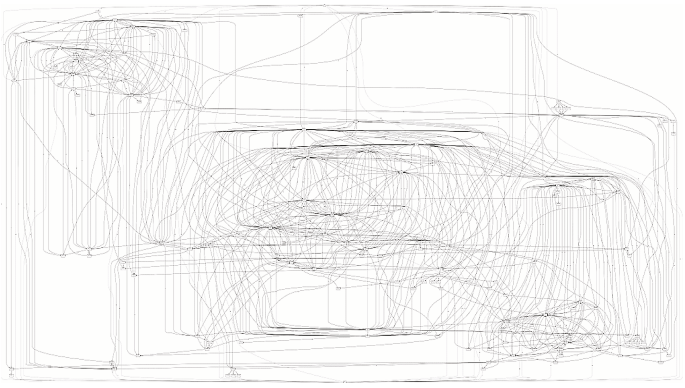


**Fig. 1.** Process model obtained by using Aperture visualising a highly complex process

To benchmark our proposed clustering we use HMM based sequential clustering. Each data sequence is described by one HMM. The distance between any pair of sequences is computed based on the loglikelihood of fitting the one sequence to the descriptive HMM of the other sequence. Once the distance matrix is built, K-means clustering is used. We use an available HMM toolbox in Matlab to develop the HMM based clustering.

The final goal of clustering business process data in this work is to improve the performance whilst predicting the data. We use Markov models and one of its extensions to predict the next process step in the experiments. The performances of these models applied to clustered data are means to evaluate the clustering strategy. In other words, they are proofs for verifying the impact of the strategy on the predictive capability.

- *MM - Markov Models*: in order to find the next task following the current one, we build transition matrices of different order Markov models.

– *MSA - Hybrid Markov Models*: a default prediction improvement module is added to higher order Markov models to obtain better accuracy. The default prediction is improved by comparing a new pattern to all the patterns from the transition matrix using sequence alignment. The most similar pattern found from the matrix is used as a substitution for the given one to contribute the prediction.

Our aim is to improve the accuracy of the prediction on the original data by clustering the data into groups of sequences with similar characteristics. Each group requires a suitable predictive model. To verify if we can cluster the data into such groups and if we can improve the accuracy by finding a suitable model for each group, we first present the MMs and MSAs performance on $DS1$ and $DS2$:

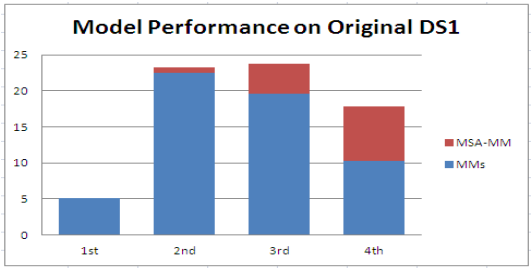



**Fig. 2.** Percentage correct of MM and MSA in predicting next process step using DS1

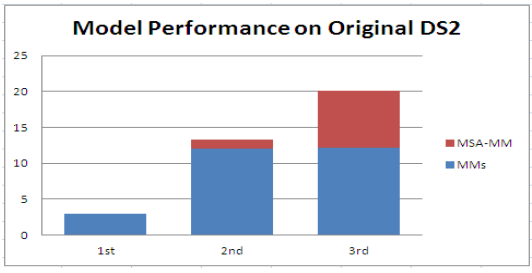



**Fig. 3.** Percentage correct of MM and MSA in predicting next process step using DS2

The results from Figures 2 and 3 show that the performances of these predictive models are quite low, only about 25%. We then introduce the performances of the same models applied to the clustered data. Both sequential methods, HMM based and our proposed clustering (SA based), are implemented to gain distance matrices. Therefore, K means clustering becomes sequential K means using such matrices as distance measure. Sequential K means HMM based and

SA based are used to cluster $DS1$, $DS2$ into 3 (or 6) and 2 clusters respectively. Figures 4 and 6 illustrate the performances of MMs and MSAs applied to 3 and 6-cluster-$DS1$, which we obtained by SA based clustering, respectively and Figures 5 and 7 illustrate the same with HMM based clustering.

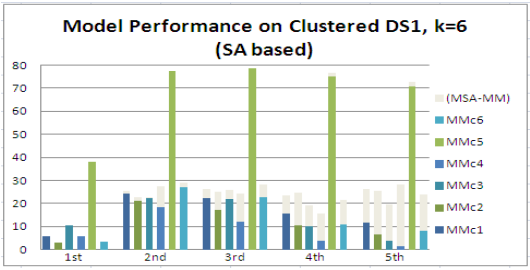



**Fig. 4.** Percentage correct of different order MMs and MSAs in predicting next process step using 6 clusters obtained by applying SA based clustering to dataset DS1

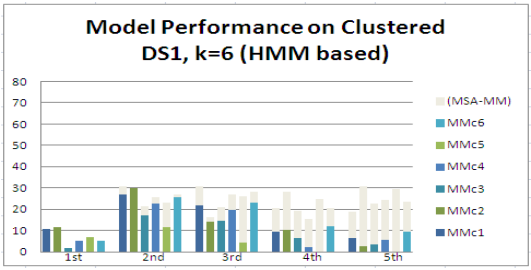



**Fig. 5.** Percentage of correct next process step predictions of different order MMs and MSAs using 6 clusters obtained by applying HMM based clustering to dataset DS1

As can be seen, in the case of K means SA based with $K = 6$, MMs and MSAs applied to cluster 5 have significantly high performance. The highest performance is 78.69 % (third order MSA) which is almost four times greater than the performance of the same models applying to the whole $DS1$ and 2.5 times comparing to these of the other clusters. Applying MMs and MSAs on clusters 4 and 6 provides better accuracy (27.27% and 28.94% respectively) than applying these on the original $DS1$ (23.76%).

In contrast, there is not much difference in terms of performance of these predictive models applying to the original data set $DS1$ or the clustered data using K means HMM based. With $K = 3$, in both cases HMM based ans SA based, there is not much change in terms of the accuracy of the process next step prediction regarding to the accuracy of the same models applied on the original $DS1$.

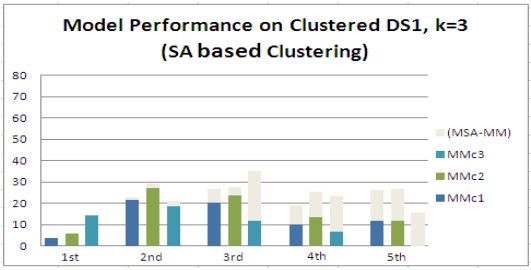



**Fig. 6.** Percentage of correct process next step predictions of different order MMs and MSAs using 3 clusters obtained by applying SA based clustering to dataset DS1

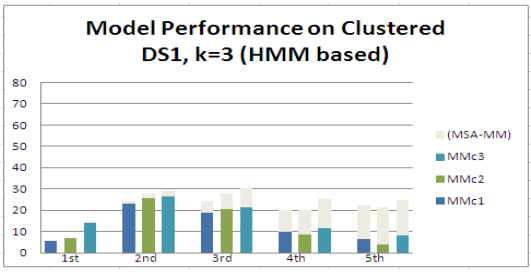



**Fig. 7.** Percentage of correct next process step predictions of different order MMs and MSAs using 3 clusters obtained by applying HMM based clustering to dataset DS1

The significant accuracy improvement in cluster 5 is the proof for our initial intuition that (1) if we manage to have subsets of data which consist of similar sequences then there exists a suitable model which performs well in each subset. Also, these models perform better in certain subsets than others and than the whole data set before being divided. (2) It indirectly proves that our proposed sequential clustering performs well, similar sequences are naturally clustered into clusters. (3) Our SA based clustering is more suitable for this type of data than the common HMM based clustering. In the case of data set $DS2$, there is not much improvement in terms of prediction accuracy after clustering the data with both clustering approaches. The highest performance of the fourth order MSA is about 27% applied to cluster 2 obtained by SA based and HMM based clusterings comparing to 20% to whole $DS2$. The performances of the first to fifth order MMs and MSAs applied to clusters 1 and 2 obtained by clustering $DS2$ in the case SA based clustering are illustrated in Figure 8.

When clustering $DS2$ using the two methods, only two clusters are formed, when we decrease the clustering goodness, more clusters are obtained but most of them have very low populations. The results of the experiments on clustered $DS1$ and $DS2$ show that different clusters need different predictive models. Higher order MMs and MSAs are especially good for data sequences in cluster 5

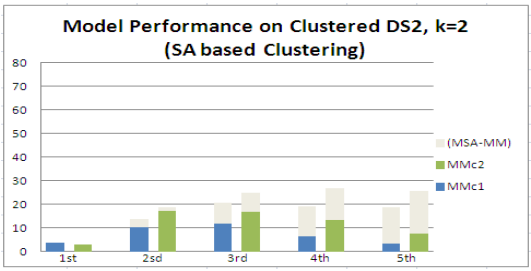



**Fig. 8.** Percentage of correct next process step predictions of different order MMs and MSAs using 2 clusters obtained by applying SA based clustering to data set DS2

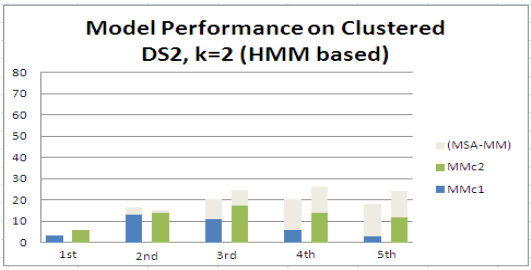



**Fig. 9.** Percentage of correct next process step predictions of different order MMs and MSAs using 2 clusters obtained by applying HMM based clustering to data set DS2

generated from $DS1$ using sequential K means clustering with $K = 6$. None of these models works well on other clusters from both data sets $DS1$ and $DS2$. It is sensible saying that the experimental models are not good for clustered data of $DS2$ as this data set is relatively small.

## 5 Conclusions

In order to deal with individual process prototypes differently, we first attempt to cluster process data into different groups of similar sequences. Such data consists of discrete symbolic sequences. After studying a number of available sequential clustering approaches, in this paper we introduce a new sequential clustering approach which is suitable for business process data. We also use the common HMM based sequential clustering in order to compare to our proposed approach. We then use predictive models to predict next process step and we significantly improve the next process step prediction in one cluster of one of the used data sets. This implies the data has been successfully clustered in a natural way and proves our strategy right.

The experimental results encourage and motivate us to continue and extend our work. The future work dicrections will explore different predictive

approaches, for example, decision trees, neural networks etc. to profit from their abilities in clustered data. We are ultimately interested in recovering the process logic even though our recover process can be the combination of a number of representative prototypes.

## References


1. Ruta, D., Majeed, B.: Business process forecasting in telecommunication industry. In: 2011 IEEE GCC Conference and Exhibition (GCC), pp. 389–392 (2011)
2. Tsui, K., Chen, V., Jiang, W., Aslandogan, Y.: Data mining methods and applications. In: Pham, H. (ed.) Handbook of Engineering Statistics, pp. 651–669. Springer (2005)
3. Trcka, N., Pechenizkiy, M.: From local patterns to global models: Towards domain driven educational process mining. In: 9th International Conference on Intelligent Systems Design and Applications, pp. 1114–1119 (2009)
4. van der Aaslt, W., Weijters, A.: Process mining: Research agenda. Computers in Industry 53(3), 231–244 (2004)
5. Smyth, P.: Clustering sequences with hidden markov models. In: Advances in Neural Information Processing Systems, pp. 648–654. MIT Press (1997)
6. Garcia, D., Parrado, E., Diaz-de Maria, F.: A new distance measure for model-based sequence clustering. IEEE Transactions on Pattern Analysis and Machine Intelligent 1(7), 1325–1331 (2009)
7. Needleman, S., Wunsch, C.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 48, 443–453 (1970)
8. Waterman, M.: Estimating statistical significance of sequence alignments. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 344, 383–390 (1994)
9. Smith, T., Waterman, M.: Identification of common molecular subsequences. Journal of Molecular Biology 147, 195–197 (1981)
10. Rajaraman, A., Ullman, J.: Mining of Massive Datasets. Cambridge University Press, Cambridge (2011)
11. Duda, R., Hart, P., Stork, D.: Pattern Classification. Wiley, New York (2001)
12. Berry, M., Linoff, G.: Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management. Wiley, Newyork (2004)
13. Gabrys, B., Bargiela, A.: General fuzzy min-max neural network for clustering and classification. IEEE Transactions on Neural Networks 11(3), 769–783 (2000)
14. Anitha Elavarasi, S., Akilandeswari, J., Sathiyabhama, B.: A survey on partition clustering algorithms. International Journal of Enterprise Computing and Business Systems 1 (2011)
15. Zaki, M., Peters, M., Assent, I., Seidl, T.: Clicks: An effective algorithm for mining subspace clusters in categorical datasets. Data Knowl. Eng. 60(1), 51–70 (2007)
16. Dhillon, S., Modha, S.: Concept decompositions for large sparse text data using clustering. Machine Learning 42, 143–175 (2001)
17. Li, C., Biswas, G.: Clustering sequence data using hidden markov model representation. In: Proceedings of the SPIE 1999 Conference on Data Mining and Knowledge Discovery: Theory, pp. 14–21 (1999)
18. Porikli, F.: Clustering variable length sequences by eigenvector decomposition using hmm. In: Fred, A., Caelli, T.M., Duin, R.P.W., Campilho, A.C., de Ridder, D. (eds.) SSPR&SPR 2004. LNCS, vol. 3138, pp. 352–360. Springer, Heidelberg (2004)

19. Kanungo, T., Mount, D., Netanyahu, N., Piatko, C., Silverman, R., Wu, A.: An efficient k-means clustering algorithm: Analysis and implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 881–892 (2002)
20. Wagstaff, K., Cardie, C., Rogers, S., Schrodl, S.: Constrained k-means clustering with background knowledge. In: 18th International Conference on Machine Learning, pp. 577–584 (2001)
21. Elkan, C.: Using the triangle ilequality to accelerate k-means. In: 20th International Conference on Machine Learning (ICML-2003), Washington DC, pp. 2–9 (2003)
22. Pham, D., Dimov, S., Nguyen, C.: Selection of k in k-means clustering. I MECH E Part C Journal of Mechanical Engineering Science 219(1), 103–119 (2005)

# A Fuzzy Semisupervised Clustering Method: Application to the Classification of Scientific Publications

Irene Diaz-Valenzuela, Maria J. Martin-Bautista, and Maria-Amparo Vila

Department of Computer Sciences and Artificial Intelligence / CITIC-UGR,
University of Granada, Spain
{idiazval,mbautis,vila}@decsai.ugr.es

**Abstract.** This paper introduces a new method of fuzzy semisupervised hierarchical clustering using fuzzy instance level constraints. It introduces the concepts of fuzzy must-link and fuzzy cannot-link constraints and use them to find the optimum $\alpha$-cut of a dendrogram. This method is used to approach the problem of classifying scientific publications in web digital libraries. It is tested on real data from that problem against classical methods and crisp semisupervised hierarchical clustering.

**Keywords:** Semisupervised clustering, Fuzzy Instance level Constraints, Digital Libraries.

## 1 Introduction

The popularity of Internet has changed the way people consult bibliographic information. This has helped to the development and widespread of digital libraries on the web. This kind of services offer information about a specific topic stored in registers like the ones in a library. It is possible to find several examples in the area of scientific publications, like DBLP, CiteSeerX, INSPEC or Google Scholar. These services make finding a specific publication very easy, but if we want to find all publications by an author it can be very hard for several reasons.

Specifically, if we consider that an author name can be written in several ways (full, abbreviated, using initials, etc.), search results will be very different according to the query made. In the same way, if there is more than one author sharing the same name, it is possible that their publications had been mixed. These problems are respectively known as *Multirepresentation* and *Identity Uncertainty*. The first one refers to the occurrence of several representations for the same name, while the second one can be defined as the impossibility of distinguish among two different authors sharing the same name.

Let us illustrate these problems with an example: looking in a digital library for an author called J. Smith. Using this query, the results can be related to James Smith, Julia Smith, Joseph Smith, among others. On the other hand, if we look for James Smith, it would not return those registers under J. Smith, even if the author uses that name in some of his works.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 179–188, 2014.

As a consequence of that, there are several data that are probably mixed and can have noise. However, it is possible to disambiguate these results and find a clean classification of publications by name using an Automatic Authority Control method [1]. That proposal was based on a hierarchical clustering algorithm combined with a proximity measure designed ad-hoc for the problem. Additionally, it needed an expert during the validation process who evaluated the solution. Even if the results of that method where quite satisfactory, the use of the expert led us to the idea of using her/his information to help the clustering process.

A good way to use external information in the process of classifying scientific publications is the use of *Semisupervised Clustering*[2]. It can be defined as a group of techniques between fully supervised classification and unsupervised clustering. These methods use a small amount of external information helping to guide the clustering process to a better solution.

The main advantage of semisupervised clustering in comparison to supervised classification is that it needs considerably less information, that it is usually provided differently that it is easier to obtain than class labels. It means that these methods can be applied to a wider range of problems. Specifically, if we focus on one of the main types of semisupervision, *instance level constraints*, it only needs information about if some pairs of instances are or are not in the same class. For that reason, these methods can be applied in situations where it is not possible or it is very hard to label the instances, as happens in the problem of classifying publications.

As previously stated, this technique fits very well the problem of classifying scientific publications by author. Given a group of publications obtained from a search in a digital library, it is not possible to know in advance the number of authors they belong to. So, as this information is not available, it is not possible to know how many groups (clusters) can be found in the data. For that reason, it is very difficult to assign them a label. However, if an expert is given two publications, she/he can say if they belong to the same author, at least with certain degree of belief.

Semisupervision has not been a very used technique in combination with hierarchical clustering. Additionally, it had always been introduced in the algorithms during the process of grouping elements. Normally, this technique had been used to drive the solution to one that fits the constraints. However, we propose to use these techniques differently. The method that we are introducing in this paper does not modify the clustering process itself but it helps to find the optimum $\alpha$-cut of the dendrogram. It allows the use of a hierarchical clustering technique that does not need to prior knowledge about how many groups will be obtained. As it was previously addressed, the problem that is been approached in this paper does not provide that information.

Instance level constraints have always been used is a crisp way, so they were limited to those problems where the expert was certain about the provided information. However, we propose to soften those conditions by the use of *fuzzy instance level constraints*. The use of these fuzzy constraints allow the expert to give a certain degree of belief about if two elements belong or not to the same cluster.

This paper is organized as follows: Section 2 gives a brief bibliography of previous work on this problem and semisupervised clustering. Section 3 introduces a formalization that will help to approach the problem of classifying publications by author. Then,

on Section 4 a new fuzzy semisupervised hierarchical clustering algorithm is proposed, introducing semisupervision as fuzzy instance level constraints. That model has been used on Section 5. Finally, some conclusions and future work are given on Section 6.

## 2 Previous Work

In the last times, Semisupervised Clustering has been a very popular area of research. Among the different variations of these techniques, *instance level constraints* as proposed by Wagstaff [3], are a very interesting model that has been receiving a lot of attention. Some of the most relevant works based on this model can be found on [2].

The problem of how to classify publication by author has been previously treated from several points of view. However, the proposed solutions typically focus on the *Identity Uncertainty* problem. It has been approached using different classification methods, both supervised and unsupervised. Among the supervised proposals, there are some that use Support Vector Machines (SVM) [4], or variations of SVM, like LASVM [5]. Other proposals include: Naive-Bayes [6], Random Forest [7], Latent Dirichlet Allocation [8] or Markov Chains [9]. On the other hand, among unsupervised methods it is possible to find proposals based on WAD method [10] or on hierarchical heuristics [11].

In contrast with previous works, in this proposal we solve at the same time the *Multirepresentation* and *Identity Uncertainty* problems. Additionally, most previous works generally use information that can not be found on some digital libraries, as author's email. The method described on this paper is focusing only on information that can be found on any digital library. An unsupervised first approach of this methodology can be found on: [1], [12]. Results from these works lead us to the development of the generic theoretical framework that has been used on this paper.

In [13] an initial proposal of how to introduce instance level constraints into hierarchical clustering was made. The use of constraints to find the optimum $\alpha$-cut of a dendrogram is different from the tendency in the literature to use them inside the clustering process itself. However, that proposal used crisp constraints, which led us to facing the idea of fuzzy instance level constraints. This new kind of constraints allow us to use them in problems where there is uncertainty or it is not possible to find a reliable expert. For that reason, this paper introduces a new method of fuzzy hierarchical semisupervised clustering based on instance level constraints. Additionally, it proposes its application to the classification of scientific publications by author.

## 3 Theoretical Framework

In this section we describe a generic theoretical framework to find a set $E = \{e_1, \ldots, e_n\}$ representing any kind of entity that can be expressed as a *set of names*. Examples of these entities can be authorities, companies, diseases, etc. Let us define:

- $N = \{n_1, \ldots, n_k\}$, a set of *names*.
- $T = \{t_1, \ldots, t_m\}$, a set of *entities* or cases, such as $\forall t \in T \,|\, t = (n, \overrightarrow{f})$; with $n \in N$, where $n$ can be in different cases and $f$ is a vector of *factors* extracted from texts.

The previous definition allow us to describe each entity as a set of names, so the problem of finding $E$ can be translated into finding a partition of the space of names: $\mathscr{P}(N) = \{N_1, \ldots, N_n\}$ such as $\forall i$; $N_i \equiv e_i$. This partition of the space of names is induced by a partition of the space of cases. In fact, it will be obtained by means of a fuzzy clustering process $\mathscr{P}(T) = (C_1, \ldots, C_n)$. So, for each $C_i = \{t_{ij} = (n_{ij}, \overrightarrow{f_{ij}})\}$, it is established that $S_i = \{n | \exists t_{ij} \in C_i; t_{ij} = (n, f_{ij})\}$. Obviously, $\cup_{i=1}^{l} S_l = N$ i.e. $\mathscr{S}(N) = \{S_i, \ldots, S_n\}$ is a cover of the space of names.

As previously stated, partition $\mathscr{P}(T)$ is obtained by means of a clustering process using a proximity measure determining the similarity among the elements to group. For that reason, it is necessary to use a measure that allow the comparison of $t$ according to the content of $N$.

Cover $\mathscr{S}(N)$ can be transformed into a partition of the space of names $\mathscr{P}(N)$ by a validation process. It involves the use of an expert and returns a set of names that are the prototypes of the partition.

It is proposed to use this model to find *authorities*, defined as the different names that can identify an author. To do that, a methodology is described using publications from Digital Libraries as *cases* and an ad-hoc proximity measure for this problem. All these concepts and methods will be detailed in the following section.

### 3.1 Authority Control in the Theoretical Framework

The previous model can be applied to solve the problem of classifying publications in a digital library by name. We can define an *authority* as a group of different representations for the name of a specific author. Using this definition, to find the set of authorities in several publications obtained from a digital library is equivalent to find the set $E$ of entities. The elements of this model can be identified in this problem as:

- Set of entities $E$ is the set of authorities.
- Set of cases $T$ is the set of publications to be classified. Each $t_i$ is a publication of the set.
- $t = (n, \overrightarrow{f})$, where $n$ is a name and $f$ is a vector of factors. Specifically, $t = (n, f_1, f_2)$ where $n$ is the author's name, $f_1$ are terms from title and abstract and $f_2$ is a list of coauthors.
- $\mathscr{P}(T)$ is a partition of the set of publications.
- $\mathscr{S}(N)$ can be obtained from the names, $n$, of $\mathscr{P}(T)$.

Digital libraries offer a wide range of information that can be used as *factors*, such as titles, authors, date of publication, venue, abstract, author's email, etc. Considering that this method it is intended to be very generic and it should be suitable for any digital library, data used as *factors* must be found in any source. Specifically it uses *terms from titles* and *authors*. Additionally, it can use *terms from abstract* if they are available. It is considered that these factors allow distinction among publication and are general enough to be found in any digital library.

Proximity measure used in this process is the opposite of a divergence measure used to compare different factors in $t$. This measures $m$ can be defined as:

$$m(t_1,t_2) = \begin{cases} 1, & \text{if } m_n = 1; \\ (1-\lambda_1-\lambda_2)\times m_n & \\ +\lambda_1 \times d_t + \lambda_2 \times m_c, & \text{otherwise.} \end{cases} \quad (1)$$

Where $m_n$ is a *name divergence measure*, $d_t$ is a *distance among terms*, $m_c$ is a *coauthors divergence measure* and $\lambda_1$, $\lambda_2$ are two weights in the interval [0,1] used to control the influence of each component. For each specific problem, $d_t$ is cosine distance. Names measure $m_n$ is obtained using a pseudo-distance specifically designed for names sharing a common componentc[12]. Coauthor's measure, $m_c$ is a modification of the previous measure for names that do not have common substring [12]. Validation of the results of this model to obtain the partition $\mathscr{P}(N)$ is made according to the process described in [1].

## 4 Fuzzy Semisupervised Clustering

In this section, a new way to introduce semisupervision into hierarchical clustering is described. It makes use of *instance level constraints*.

Hierarchical clustering is a family of clustering algorithms that instead of creating a partition of the input dataset, it creates a hierarchy of elements in a binary tree called dendrogram. Each dendrogram is equivalent to an ultrametric distance matrix $\mathscr{U}$ and generates a fuzzy similarity matrix for the elements to be grouped [14]. This kind of clustering does not generate a partition of the input elements but a set of nested partitions according to several levels $\alpha$. Those $\alpha$ are just the $\alpha$-cuts of the fuzzy similarity matrix. Then, the problem of finding the best partition of the dendrogram to obtain the best partition of the input elements is equivalent to the problem of finding the best $\alpha-$cut of the fuzzy similarity matrix. There are objective unsupervised methods to find that partition like the one on [15].

On previous works we have used the unsupervised method to find the optimum $\alpha$-cut of the partition without using external information [1], [12]. Now, it is our goal to find the optimum partition using *instance level constraints* [3]. These constraints are given by an expert with knowledge about some of input data.

Instance level constraints indicate if two elements are or are not in the same group. It is possible to distinguish two kinds: *must-link* and *cannot-link*. Let us define them as:

- *MUST-LINK* $\mathscr{M}$: Given $t_i$ and $t_j$, if $\exists$ a $\mathscr{M}(t_i,t_j)$, then $t_i$ and $t_j$ must be in the same cluster.
- *CANNOT-LINK* $\mathscr{H}$: Given $t_i$ and $t_j$, if $\exists$ a $\mathscr{H}(t_i,t_j)$, then $t_i$ and $t_j$ cannot not be in the same cluster.

Traditionally, this kind of constraints have been given in a crisp way. However, crisp information is not appropriate for problems where there is uncertainty or it is not possible to find an expert that can give true knowledge about the working data, or even, there are two experts that do not agree in their information. In these cases, the use of fuzzy

constraints would allow us to soften the conditions by giving a degree of belief about if two elements are or are not in the same group. With these ideas, the previous definition can be redefined as:

- *FUZZY MUST-LINK* $\mathscr{M} = \beta$: Given $t_i$ and $t_j$, if $\exists$ a $\mathscr{M}(t_i,t_j) = \beta$, then $t_i$ and $t_j$ belong to the same cluster with degree $\beta \in [0,1]$.
- *FUZZY CANNOT-LINK* $\mathscr{H} = \gamma$: Given $t_i$ and $t_j$, if $\exists$ a $\mathscr{H}(t_i,t_j)$, then $t_i$ and $t_j$ are not in the same cluster with $\gamma \in [0,1]$.

It is possible to represent the constraints between pairs of elements using two matrices $M_m$ and $M_c$ such as:

$$M_m(t_i,t_j) = \begin{cases} \beta, & \text{if} \exists \mathscr{M}(t_i,t_j) \\ -1, & \text{if} \nexists \mathscr{M}(t_i,t_j) \end{cases} \tag{2}$$

$$M_c(t_i,t_j) = \begin{cases} \gamma, & \text{if} \exists \mathscr{H}(t_i,t_j) \\ -1, & \text{if} \nexists \mathscr{H}(t_i,t_j) \end{cases} \tag{3}$$

Such as $\nexists(t_i,t_j) \in T | \mathscr{M}(t_i,t_j) = \beta$ and $\mathscr{H}(t_i,t_j) = \gamma$, meaning that there can not be two restrictions *must-link* and *cannot-link* for the same pair of elements.

In the same way, for each $\alpha$-cut of the dendrogram it is also possible to create a matrix $C$ from the opposite of the matrix $U$ representing the elements that have been clustered on each step. So, for each $\alpha$-cut there is a matrix $C_\alpha$ with the elements joined at that specific $\alpha$-level. Each $C_\alpha(t_i,t_j)$ will have two possible values:

- $C_\alpha(t_i,t_j) = 1$ if elements $t_i$ and $t_j$ are in the same group
- $C_\alpha(t_i,t_j) = 0$ if elements $t_i$ and $t_j$ are not in the same group

Matrices $M_m$ and $M_c$ can be compared with the different matrices $C_\alpha$ according to (4).

$$s(M(t_i,t_j), C_\alpha(t_i,t_j)) = \begin{cases} v, & \text{if } M_m(t_i,t_j) \geq \alpha \text{ and } C_\alpha(t_i,t_j) = 1; \\ -w, & \text{if } M_m(t_i,t_j) \geq \alpha \text{ and } C_\alpha(t_i,t_j) = 0; \\ v, & \text{if } M_c(t_i,t_j) \geq \alpha \text{ and } C_\alpha(t_i,t_j) = 0; \\ -w, & \text{if } M_c(t_i,t_j) \geq \alpha \text{ and } C_\alpha(t_i,t_j) = 1; \\ 0, & \text{otherwise .} \end{cases} \tag{4}$$

Where $v$ is used to adjust the weight that reinforce the constraint have in the expression and $w$ is the penalty of contradict the constraint. For this problem, we suggest to use $v = 1$ and $w = 0.5$, but this values can be changed according to the problem or the confidence of the expert.

So, the optimum $\alpha$-cut can be calculated by maximizing the expression in (5)

$$S = \max_{\alpha \in [0,1]} \{ \sum_{i,j}^{n} s(M(t_i,t_j), C_\alpha(t_i,t_j)) \} \tag{5}$$

The $\alpha$ from matrix $C_\alpha$ for which (5) have its maximum is the optimum $\alpha$-cut of the dendrogram. Cutting the tree at that level will give us the optimum partition of the dataset.

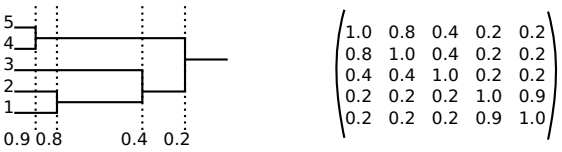



**Fig. 1.** Example of dendrogram and its equivalent opposite of the ultrametric matrix

### 4.1 Example

Let us illustrate this method with an example. Figure 1 shows an example dendrogram and its equivalent opposite of the ultrametric matrix returned by a clustering process over 5 elements. An expert with knowledge about the problem has given the following constraints:

- $M_m(1,2) = 0.8$
- $M_m(4,5) = 1.0$
- $M_c(2,3) = 0.9$
- $M_c(1,5) = 0.6$

From the ultrametric matrix we have four different $\alpha$-cuts: $\alpha = \{0.9, 0.8, 0.4, 0.2\}$. For each one of them we are going to evaluate $s(M, C_\alpha)$, with values of $v = 1$, $w = 0.5$.

So, for each $\alpha$:

$$s(M(t_i,t_j),C_{0.9}) = \begin{cases} M_m(1,2) = 0.8 < 0.9 \text{ and } C_{0.9}(1,2) = 0;\ 0 \\ M_m(4,5) = 1.0 \geq 0.9 \text{ and } C_{0.9}(4,5) = 1;\ 1 \\ M_c(2,3) = 0.9 \geq 0.9 \text{ and } C_{0.9}(2,3) = 0;\ 1 \\ M_c(1,5) = 0.6 < 0.9 \text{ and } C_{0.9}(1,5) = 0;\ 0 \end{cases} = 2 \tag{6}$$

$$s(M(t_i,t_j),C_{0.8}) = \begin{cases} M_m(1,2) = 0.8 \geq 0.8 \text{ and } C_{0.8}(1,2) = 1;\ 1 \\ M_m(4,5) = 1.0 \geq 0.8 \text{ and } C_{0.8}(4,5) = 1;\ 1 \\ M_c(2,3) = 0.9 \geq 0.8 \text{ and } C_{0.8}(2,3) = 0;\ 1 \\ M_c(1,5) = 0.6 < 0.8 \text{ and } C_{0.8}(1,5) = 0;\ 0 \end{cases} = 3 \tag{7}$$

$$s(M(t_i,t_j),C_{0.4}) = \begin{cases} M_m(1,2) = 0.8 \geq 0.4 \text{ and } C_{0.4}(1,2) = 1;\ 1 \\ M_m(4,5) = 1.0 \geq 0.4 \text{ and } C_{0.4}(4,5) = 1;\ 1 \\ M_c(2,3) = 0.9 \geq 0.4 \text{ and } C_{0.4}(2,3) = 1;\ -0.5 \\ M_c(1,5) = 0.6 \geq 0.4 \text{ and } C_{0.4}(1,5) = 0;\ 0 \end{cases} = 1.5 \tag{8}$$

$$s(M(t_i,t_j),C_{0.2}) = \begin{cases} M_m(1,2) = 0.8 \geq 0.2 \text{ and } C_{0.2}(1,2) = 1;\ 1 \\ M_m(4,5) = 1.0 \geq 0.2 \text{ and } C_{0.2}(4,5) = 1;\ 1 \\ M_c(2,3) = 0.9 \geq 0.2 \text{ and } C_{0.2}(2,3) = 1;\ -0.5 \\ M_c(1,5) = 0.6 \geq 0.2 \text{ and } C_{0.2}(1,5) = 1;\ -0.5 \end{cases} = 1 \tag{9}$$

Now, using (5) we get:

$$S = \max_{\alpha \in [0,1]} s(M, C_\alpha) = \max\{2, 3, 1.5, 1\} = 3 \rightarrow \alpha = 0.8 \tag{10}$$

So, the optimum $\alpha$-cut is found at $\alpha = 0.8$ obtaining three clusters.

## 5 Experimental Results

This section will show the behaviour of fuzzy instance level constraints to obtain the optimum $\alpha$-cut of a dendrogram. They will be compared with crisp instance level constraints[12] and classic unsupervised methods[15].

This method has been tested using two datasets obtained from real searches of the surname Smith over *DBLP* and *CiteSeerX*. Resulting datasets contain 500 and 485 instances respectively. It is our intention to find out how many instance level constraints this method needs to find the optimum $\alpha$-cut properly. For that reason, these data had

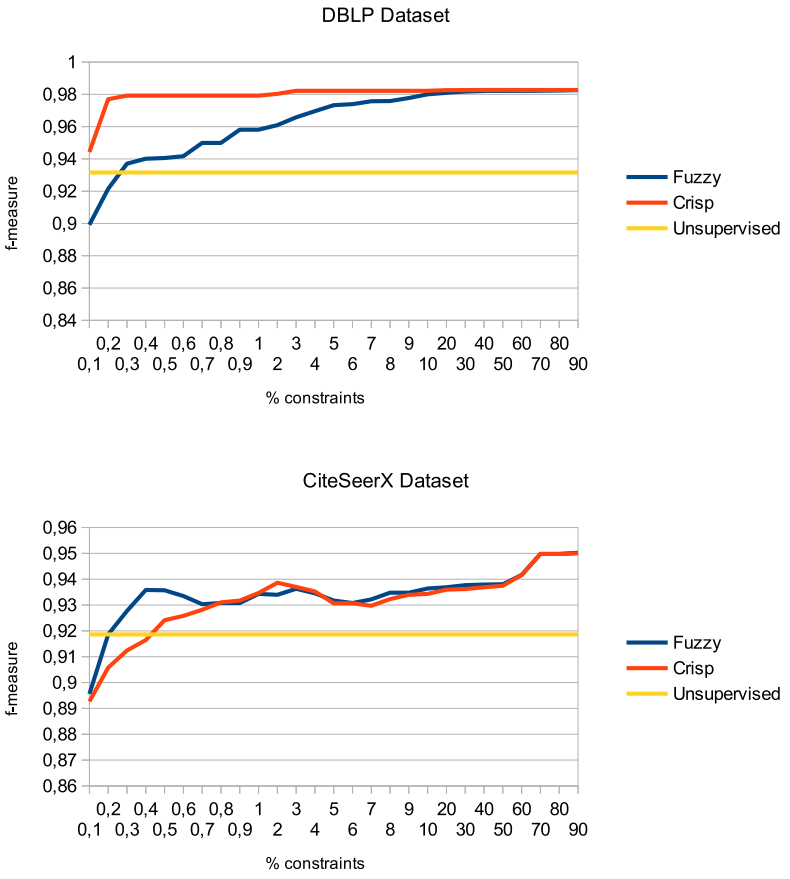



**Fig. 2.** Comparison of the different methods to obtain the $\alpha$-cut of a dengrogram using data extracted from DBLP and CiteSeerX. Horizontal axis show the percentage of constraints used on the method. Line named Fuzzy correspond with the method using fuzzy instance level constraints, line named crisp correspond to crisp instance level constraints and line named Unsupervised correspond with the unsupervised method to obtain the optimum $\alpha$-cut of the dendrogram.

been handed over to an expert who had given us fuzzy instance level constraints for all pair of instances. It is important to point out that all constraints with a degree of belief lower than 0.5 have been discarded, as information under this level is not considered reliable.

Previously described data had gone through a fuzzy clustering algorithm with the divergence measure described on [1]. Resulting dendrogram is cut using three different methods: a classic unsupervised one described on [15], crisp instance level constraints [13] and fuzzy instance level constraints. For each of the methods using instance level constraints, the cut had been performed with a different amount of constraints that is increased iteratively. Starting with a very small amount of constraints, 0.1%, they have been increased up to 90%. For each of the $\alpha$-cuts of the dendrogram, the f-measure [16] of the resulting clusters have been obtained. Results from this experiment can be seen on Figure 2.

On the Figure it is possible to observe a very similar behaviour. The lines corresponding to the fuzzy and crisp method outperform the unsupervised method using a small amount of constraints. It means that the use of instance level constraints help to find the optimum $\alpha$-cut of a dendrogram. However, if we compare the fuzzy and crisp method, we can see that even if the crisp method needs less constraints to work properly, the similarity between the crisp and fuzzy problems indicate that it is possible use fuzzy semisupervised clustering on problems with uncertainty or unreliable experts.

Additionally, as it is possible to see in the Figure, the results tend to stabilize after a determinate amount of restrictions. It indicates that the expert only need to provide information about between 5% or 10% of the pairs of instances. It means that the number of constraints needed for a proper performance of this method is quite low, which make easier to find experts that can be used on these problems.

## 6 Conclusions

In this paper the use of fuzzy instance level constraints on hierarchical semisupervised clustering had been studied. These constraints have the advantage against crisp constraints that they can be used in problems where the expert cannot provide certain information.

At it had been proved, its use on the problem of the classification of scientific publications outperform classical methods. Additionally, its behaviour is very similar to crisp instance level constraints which allow its application on those cases where the expert cannot give certain data.

It remains as a future work the study of the entropy of the constraints given by the expert, as well as the study of the influence of the distinct types of constraints in the final solution. Another very interesting line of work could be the use of these constraints on other kinds of clustering or use them not only on the process of finding the optimum $\alpha$-cut of a dendrogram, but on the clustering itself.

**Acknowledgments.** This research has been partially funded by Junta de Andalucia projects P10-TIC6109 and P11-TIC7460. Furthermore, this work has been supported by the Spanish Ministry of Education by “Programa de Formacion del Profesorado Universitario (FPU)”.

## References


1. Diaz-Valenzuela, I., Martin-Bautista, M.J., Vila, M.A., Campaña, J.R.: An automatic system for identifying authorities in digital libraries. Expert Systems with Applications 40, 3994–4002 (2013)
2. Davidson, I., Basu, S.: A survey of clustering with instance level constraints. ACM Transactions on Knowledge Discovery from Data, 1–41 (2007)
3. Wagstaff, K., Cardie, C.: Clustering with instance-level constraints. In: Proceedings of the Seventeenth International Conference on Machine Learning, pp. 1103–1110 (2000)
4. Peng, H.T., Lu, C.Y., Hsu, W., Ho, J.M.: Disambiguating authors in citations on the web and authorship correlations. Expert Systems with Applications 39, 10521–10532 (2012)
5. Huang, J., Ertekin, S., Giles, C.L.: Efficient name disambiguation for large-scale databases. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) PKDD 2006. LNCS (LNAI), vol. 4213, pp. 536–544. Springer, Heidelberg (2006)
6. Han, H., Giles, L., Zha, H., Li, C., Tsioutsiouliklis, K.: Two supervised learning approaches for name disambiguation in author citations. In: Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, pp. 296–305 (2004)
7. Treeratpituk, P., Giles, C.L.: Disambiguating authors in academic publications using random forests. In: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2009, pp. 39–48. ACM, Austin (2009)
8. Ferreira, A.A., Gonçalves, M.A., Almeida, J.M., Laender, A.H.F., Veloso, A.: A tool for generating synthetic authorship records for evaluating author name disambiguation methods. Information Sciences 206, 42–62 (2012)
9. Tang, J., Fong, A.C.M., Wang, B., Zhang, J.: A unified probabilistic framework for name disambiguation in digital library. IEEE Trans. on Knowl. and Data Eng. 24, 975–987 (2012)
10. Pereira, D.A., Ribeiro-Neto, B., Ziviani, N., Laender, A.H.: Using web information for author name disambiguation. In: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2009, pp. 49–58. ACM, Austin (2009)
11. Cota, R.G., Ferreira, A.A., Nascimento, C., Gonçalves, M.A., Laender, A.H.F.: An unsupervised heuristic-based hierarchical method for name disambiguation in bibliographic citations. Journal of the American Society for Information Science and Technology 61, 1853–1870 (2010)
12. Diaz-Valenzuela, I., Martin-Bautista, M.J., Vila, M.A.: A proposal for automatic authority control in digital libraries. Information Processing and Magnament (2013)
13. Diaz-Valenzuela, I., Martin-Bautista, M.J., Vila, M.A.: Using a semisupervised fuzzy clustering process for identity identification in digital libraries. In: IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint, pp. 831–836 (2013)
14. Kaufmann, A.: Introduction to the Theory of Fuzzy Subsets. Academic Pr. (1975)
15. Delgado, M., Gomez-Skarmeta, A.F., Vila, M.A.: On the use of hierarchical clustering in fuzzy modeling. International Journal of Approximate Reasoning 14, 237–257 (1996)
16. Baeza-Yates, R.A., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley Longman Publishing Co., Inc., Boston (1999)

# Using a Fuzzy Based Pseudometric in Classification

Sandra Sandri, Flávia Martins-Bedê, and Luciano Dutra

Instituto Nacional de Pesquisas Espaciais
12201-970, São José dos Campos, SP
sandra.sandri@inpe.br, {flavinha,luciano.dutra}@dpi.inpe.br

**Abstract.** In this work, we propose a pseudometric based on a fuzzy relation, which is itself derived from a fuzzy partition. This pseudometric is a metric in the particular case in which the fuzzy partition is composed solely by triangular fuzzy sets. We prove that these functions are indeed a pseudometric and a metric and illustrate their use in an experiment for the classification of land use in an area of the Brazilian Amazon region.

**Keywords:** classification, k-NN, fuzzy partitions, distance, metric, pseudometric.

## 1 Introduction

One of the main approaches to classification is the so-called k-NN classifiers, in which an element of a domain is assigned the class that represents the majority of the classes of its closest k neighbours (see [9] for a review of classification techniques). The function used to obtain the neighbors of an element $x$ in a multi-dimensional domain $O$ is usually a metric (or distance) on $O$, but there are a few works proposing the use of pseudometrics in general instead (see for instance [4]).

In this work, we propose the use of a function that is the complement in $[0, 1]$ of a particular kind of similarity relation, called an Order Compatible Fuzzy Relation (OCFR$_\preceq$), defined using a total order $(\Omega, \preceq)$ [8]. An OCFR$_\preceq$ itself is derived from a particular type of fuzzy partition (a collection of fuzzy sets), called Convex Fuzzy Partitions (CFP$_\preceq$). The creation of OCFR$_\preceq$ was motivated by the need to ease the burden of creating suitable relations for use in a particular fuzzy case-based reasoning classification approach [6].

The main goal of this work is to prove that the proposed function is i) a pseudometric, when obtained from a specific type of CFP$_\preceq$, called 2-Ruspini, and, in particular, ii) a metric, when this CFP$_\preceq$ is moreover composed solely of triangular fuzzy sets. We also briefly describe an application of this function in the classification of land use in an area of the Brazilian Amazon region.

## 2 Fuzzy Convex Partitions and Order-Compatible Relations

In Fuzzy Sets Theory, membership to sets is no longer an all-or-nothing notion [1]. A fuzzy set $A$ on a domain $\Omega$ is characterized by a mapping $A : \Omega \rightarrow [0, 1]$, called the membership function of $A$. It is said to be normalized when $\exists\ x_0 \in \Omega$ such that

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 189–198, 2014.

$A(x_0) = 1$. A level cut of $A$ is defined as $\forall \alpha \in (0,1], [A]_\alpha = \{x \in \Omega \mid A(x) \geq \alpha\}$. The core and support of a fuzzy set $A$ are particular types of level cuts defined as $core(A) = \{x \in \Omega \mid A(x) = 1\}$ and $supp(A) = \{x \in \Omega \mid A(x) > 0\}$), respectively. Particular types of fuzzy sets are those described by linear by parts membership functions, in which all level cuts are (nested) intervals of the domain. If $A_0$ is such a fuzzy set and its core is given by $\exists\ x_l, x_u \in \Omega, core(A_0) = [x_l, x_u]$, then $A_0$ is trapezoidal (respec. triangular), when $x_l \neq x_u$ (respec. $x_l = x_u$). A collection of fuzzy sets is usually called a fuzzy partition, with more specific definitions depending on the properties obeyed by the fuzzy sets composing the partition.

A fuzzy relation $S$ is characterized by a mapping from a multidimensional domain $O = \Omega_1 \times ... \times \Omega_n$ to $[0,1]$. Its normalization is defined similarly to that for fuzzy sets in one-dimensional domains. Given two distinct fuzzy relations $S$ and $S'$, we say that $S$ is *finer* than $S'$, when $\forall x, y \in \Omega, S(x,y) \leq S'(x,y)$.

A $T$-norm operator is a mapping $\top : [0,1]^2 \rightarrow [0,1]$, that is commutative, associative, monotonic and has 1 as neutral element. Given a $T$-norm $\top$, its associated residuated implication operator $I_\top$ is defined as [7].

$$I_\top(x,y) = \sup\{z \in [0,1] \mid \top(x,z) \leq y\},$$

and its associated biresiduation $BI_\top$ is defined by

$$BI_\top(x,y) = \min(I_\top(x,y), I_\top(y,x)).$$

In particular, the Lukasiewicz $T$-norm operator is defined as

$$\top_L(x,y) = \max(x + y - 1, 0).$$

and its associated residuated operator and biresiduation are respectively defined as

- $I_L(x,y) = \min(1 - x + y, 1)$,
- $BI_L(x,y) = 1 - \mid x - y \mid$,

where $\mid . \mid$ denotes the absolute value of a real number.

Let $S : \Omega^2 \rightarrow [0,1]$ be a fuzzy binary relation and $(\Omega, \preceq)$ be a total order. Formally, $S$ is an Order Compatible Fuzzy Relation with Respect to a Total Order $(\Omega, \preceq)$ (*OCFR*$_\preceq$ or *OCFR*, for short), when it obeys the following properties [8]:

- $\forall x, y, z \in \Omega$, $S(x,x) = 1$ (*reflexivity*)
- $\forall x, y, z \in \Omega$, $S(x,y) = S(y,x)$ (*symmetry*)
- $\forall x, y, z \in \Omega$, if $x \preceq y \preceq z$, then $S(x,z) \leq \min(S(x,y), S(y,z))$ (*compatibility with total order* $(\Omega, \preceq)$, or $\preceq$*-compatibility* for short).

Let $(\Omega, \preceq)$ be a total order and let $\mathbf{A} = \{A_1, ..., A_t\}$ be a collection of fuzzy sets in $\Omega$. Formally, $\mathbf{A}$ is a Convex Fuzzy Partition with Respect to a Total Order $(\Omega, \preceq)$ (*CFP*$_\preceq$ or *CFP*, for short), if it obeys the following properties [8]:

1. $\forall A_i \in \mathbf{A}, \exists x \in \Omega, A_i(x) = 1$ (*normalization*),
2. $\forall x, y, z \in \Omega, \forall A_i \in \mathbf{A}$, if $x \preceq y \preceq z$ then $A_i(y) \geq \min(A_i(x), A_i(z))$ (*convexity*),

3. $\forall x \in \Omega, \exists A_i \in \mathbf{A}, A_i(x) > 0$ (*domain-covering*),
4. $\forall A_i, A_j \in \mathbf{A}$, if $i \neq j$ then $core(A_i) \cap core(A_j) = \emptyset$ (*non-core-intersection*).

Let $\mathcal{A}_{(\Omega,\preceq)}$ denote the set of all CFPs that can be derived considering a total order $(\Omega, \preceq)$. CFP $\mathbf{A} \in \mathcal{A}_{(\Omega,\preceq)}$ is said to be a *$n$-CFP* if each element in $\Omega$ has non-null membership to at most $n$ fuzzy sets in $\mathbf{A}$ ($n \geq 1$). In particular, a 2-CFP$_\preceq$ $\mathbf{A}$ is called a *2-Ruspini* partition, when it obeys additivity:

– $\forall x \in \Omega, \sum_i A_i(x) = 1$ (*additivity*)

In [8], the authors propose to generate OCFR$_\preceq$ $S^+ : \Omega^2 \to [0, 1]$ from a CFP$_\preceq$ $\mathbf{A}$ as

$$S^+(x, y) = \begin{cases} 0, \text{if } S^*(x, y) = 0 \\ S_L(x, y), \text{otherwise} \end{cases}$$

$$\forall x, y \in \Omega, S^*(x, y) = \sup_i \min(A_i(x), A_i(y))$$

$$\forall x, y \in \Omega, S_L(x, y) = inf_i \;\; 1- \mid A_i(x) - A_i(y) \mid$$

Note that $S_L$ is constructed based on the Lukasiewicz biresiduated operator $BI_L$.

When $A$ is a 2-Ruspini partition, $S^+(x, y)$ obeys the following properties (see [8] for proofs):

– $\forall A_i \in \mathbf{A}, \forall c \in core(A_i), \forall x \in \Omega, S(c, x) = S(x, c) = A_i(x)$ (*core-restrictivity*);
– $\forall x, y \in \Omega, S(x, y) \geq S^*(x, y) = \sup_i \min(A_i(x), A_i(y))$ (*level-cut-compatibility*).

Core-restrictivity states that the "slice" from $S$, corresponding to an element $c$ at the core of a set $A_i$ (the column of element $c$ in [7]), is exactly the same as $A_i$, whereas level-cut-compatibility ensures that any two elements of $\Omega$ that belong to level cut $[A_i]_\alpha$ of a fuzzy set $A_i$ in $\mathbf{A}$ will also belong to level cut $[S]_\alpha$ of relation $S$ [8].

## 3 Function $f^+$

A metric, or distance function, $d : \Omega \to \mathbb{R}$ satisfies the following properties:

– $\forall x, y \in \Omega$, $d(x, y) \geq 0$ (non-negativity)
– $\forall x, y \in \Omega$, $d(x, y) = 0$ if and only if $x = y$ (identity of indiscernibles)
– $\forall x, y \in \Omega$, $d(x, y) = d(y, x)$ (symmetry)
– $\forall x, y \in \Omega$, $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

A pseudometric satisfies non-negativity, symmetry and the triangle inequality, but the identity of indiscernibles property is substituted by a weaker property:

– $\forall x \in \Omega, d(x, x) = 0$ (anti-reflexivity)

Let $\mathbf{A}$ be a CFP$_\preceq$ and $S^+_\mathbf{A}$ be the result of applying $S^+$ to $\mathbf{A}$. Here we propose the use of function $f^+_\mathbf{A} : \Omega^2 \in [0, 1]$ in tasks in which metrics and pseudometrics are employed:

$$\forall x, y \in \Omega, f^+_\mathbf{A}(x, y) = 1 - S^+_\mathbf{A}(x, y).$$

This formula can be written directly as:

$$\forall x, y \in \Omega, f^+_\mathbf{A}(x, y) = \begin{cases} 1, \text{if } \forall i, \min(A_i(x), A_i(y)) = 0, \\ \sup_i \ \mid A_i(x) - A_i(y) \mid, \text{otherwise.} \end{cases}$$

When no confusion is possible, we denote $f^+_\mathbf{A}$ as simply $f^+$.

Below, we first prove some properties of $f^+$ and then prove that it is a pseudometric in general and a metric when $\mathbf{A}$ is composed solely of triangular fuzzy sets. Note that, since $S^+_\mathbf{A}$ obeys the compatibility with total order property, we obtain the inequality

$$\forall x, y \in \Omega, f^+_\mathbf{A}(x, z) \geq \max(f^+_\mathbf{A}(x, y), f^+_\mathbf{A}(y, z)).$$

Therefore, that property leads us to obtain a lower bound for $f^+_\mathbf{A}(x, z)$, in terms of $f^+_\mathbf{A}(x, y)$ and $f^+_\mathbf{A}(y, z)$, when $x \preceq y \preceq z$. By proving the triangle inequality for $f^+_\mathbf{A}$, we will then obtain an upper bound for $f^+_\mathbf{A}$, that does not depend on the order of $x, y$ and $z$ in $\Omega$.

In the following, we say that two elements $p$ and $q$ in $\Omega$ *relate to each other with respect to a* CFP$_\preceq$ $\mathbf{A}$, when they both have non-null membership degree to at least one fuzzy set in $\mathbf{A}$, i.e when $\exists A \in \mathbf{A}, \min(A(p), A(q)) > 0$.

Let $\mathbf{A}$ be a 2-Ruspini CFP$_\preceq$. Each element in $\Omega$ has non-null membership to either 1 or 2 fuzzy sets in $\mathbf{A}$, due to the additivity and covering properties. Therefore, with respect to $\mathbf{A}$, any two elements $p$ and $q$ in $\Omega$ can either: i) be unrelated, when $\forall A_i \in \mathbf{A}$, $\min(A_i(p), A_i(q)) = 0$, ii) be related by a single fuzzy set, when $\exists B \in \mathbf{A}$ such that $p, q \in supp(B)$ and $\forall A_i \in \mathbf{A}$, $A_i \neq B$, $\min(A_i(p), A_i(q)) = 0$, or iii) be related by exactly two fuzzy sets, when there exists two contiguous fuzzy sets $B_1, B_2$ in $\mathbf{A}$ such that $p, q \in supp(B_1) \cap supp(B_2)$ and for all $A_i \in \mathbf{A}$ such that $A_i \neq B_1$ and $A_i \neq B_2$, $\min(A_i(p), A_i(q)) = 0$.

**Corollary 1.** *Let $(\Omega, \preceq)$ be a total order, $\mathbf{A}$ be a 2-Ruspini CFP$_\preceq$, $S^+(x, y)$ be an OCFR derived from $\mathbf{A}$ and $f^+ = 1 - S^+$.*

1. *If $p$ and $q$ are unrelated, we have $f^+(p, q) = 1$.*

   *Proof. In this case, for all $A_i \in \mathbf{A}$, $\min(A_i(p), A_i(q)) = 0$, and thus the result follows from the definition of $f^+$.*

2. *If $p$ and $q$ are related by a single fuzzy set $B \in \mathbf{A}$, then either (a) $f^+(p, q) = 0$, when $p$ and $q$ both belong to the core of $B$, or (b) $f^+(p, q) = \max(1 - B(p), 1 - B(q))$, otherwise.*

   *Proof. In this case, there exists $B \in \mathbf{A}$ such that $p, q \in supp(B)$ and for all $A_i \in \mathbf{A}$, $A_i \neq B$, $\min(A_i(p), A_i(q)) = 0$.*

- *a) If $p$ and $q$ belong to the core of $B$, $B(p) = B(q) = 1$ and thus $\mid B(p) - B(q) \mid = 0$. Due to additivity in 2-Ruspini partitions, for all $A_i \in \mathbf{A}$ such that $A_i \neq B$, $\min(A_i(p), A_i(q)) = 0 = \mid A_i(p) - A_i(q) \mid$. Therefore, $f^+(x, z) = 0$.*

- *b) If $p$ and $q$ do not both belong to the core of $B$, then there exist two fuzzy sets $B^-$ and $B^+$ in $\mathbf{A}$ adjacent to $B$, to its left and right, respectively, which are such that $p \in supp(B^-) \cap supp(B)$ and $q \in supp(B) \cap supp(B^+)$. All fuzzy sets $A_i \in \mathbf{A}$ such that $A_i \notin \{B^-, B, B^+\}$ can be disregarded in the calculation of $f^+(p, q)$, because for any $A$ in $\mathbf{A}$, if $\min(A(p), A(q)) = 0$, then $\mid A(p) - A(q) \mid = 0$. We thus have $\forall p, q \in supp(B)$, $f^+(p, q) = \max_{A \in \{B^-, B, B^+\}} \mid A(p) - A(q) \mid = \max(\mid B^-(p) - B^-(q) \mid, \mid B(p) - B(q) \mid, \mid B^+(p) - B^+(q) \mid)$. But $B^+(p) = B^-(q) = 0$, because $p \notin supp(B^+)$ and $q \notin supp(B^-)$. Moreover, $\forall p \in supp(B^-) \cap supp(B^+)$, we have $B^-(p) = 1 - B(p)$ and $\forall q \in supp(B) \cap supp(B^+)$, $B^+(q) = 1 - B(q)$, due to additivity in $\mathbf{A}$. We thus have $f^+(p, q) = \max(\mid 1 - B(p) - 0 \mid, \mid B(p) - B(q) \mid, \mid 0 - 1 + B(q) \mid) = \max(1 - B(p), 1 - B(q))$.*

3. *If $p$ and $q$ are related by 2 fuzzy sets $B_1$ and $B_2$ in $\mathbf{A}$, we have $f^+(p, q) = \mid B_1(p) - B_1(q) \mid = \mid B_2(p) - B_2(q) \mid$.*

   *Proof. In this case, there exists two contiguous fuzzy sets $B_1, B_2$ in $\mathbf{A}$ such that $p, q \in supp(B_1) \cap supp(B_2)$ and for all $A_i \in \mathbf{A}$ such that $A_i \neq B_1$ and $A_i \neq B_2$, $\min(A_i(p), A_i(q)) = 0$. As seen previously, we only have to consider fuzzy sets $B_1$ and $B_2$ from $\mathbf{A}$. We thus have $f^+(p, q) = \max(\mid B_1(p) - B_1(q) \mid, \mid B_2(p) - B_2(q) \mid)$. But since $\mathbf{A}$ is 2-Ruspini and $B_1$ and $B_2$ are contiguous, $f^+(p, q) = \mid B_1(p) - B_1(q) \mid = \mid B_2(p) - B_2(q) \mid$, which completes the proof.*

**Corollary 2.** *Let $(\Omega, \preceq)$ be a total order and $\mathbf{A}$ be a 2-Ruspini CFP$_\preceq$ derived from $\mathbf{A}$. Function $f^+$ derived from $\mathbf{A}$ satisfies the triangle inequality.*

*Proof.* To prove triangle inequality, we have to verify that

$$\forall x, y, z \in \Omega,\ f^+(x, z) \leq f^+(x, y) + f^+(y, z).$$

Let $supp_p$ stand for the union of supports of the fuzzy sets to which an element $p$ in $\Omega$ has non-empty membership, i.e. $\forall p \in \Omega$, $supp_p = \cup_i \{supp(A_i) \mid A_i(p) > 0\}$. We have to verify two cases.

- Case 1: $x$ and $z$ are unrelated in $\mathbf{A}$, i.e. for all $A_i \in \mathbf{A}$, $\min(A_i(x), A_i(z)) = 0$.

  In this case, by definition, $f^+(x, z) = 1$. When $y \notin supp_x \cup supp_z$, i.e. $y$ is related to neither $x$ nor $z$ in $\mathbf{A}$, we have $f^+(x, y) = f^+(y, z) = 1$ and thus $f^+(x, z) = 1 \leq f^+(x, y) + f^+(y, z) = 2$. When $y \in supp_x$, we have $f^+(y, z) = 1$, since by hypothesis $supp_x \cap supp_z = \emptyset$, and therefore $f^+(x, z) = 1 \leq f^+(x, y) + f^+(y, z) = f^+(x, y) + 1$. We use a similar reasoning when $y \in supp_z$.

- Case 2: $x$ and $z$ are related in $\mathbf{A}$, i.e. $\exists A \in \mathbf{A}$, $\min(A(x), A(z)) > 0$.

  When $y \notin supp_x \cup supp_z$, i.e. $y$ is related to neither $x$ nor $z$ in $\mathbf{A}$, we have $f^+(x, y) = f^+(y, z) = 1$. We thus obtain $f^+(x, z) \leq f^+(x, y) + f^+(y, z) = 2$.

Let us now examine the case in which $y$ is related to either $x$ or $z$ or both. As mentioned above, since $\mathbf{A}$ is a 2-Ruspini CFP$_{\preceq}$, we have two possibilities for $x$ and $z$; they are related by either 1 or 2 fuzzy sets.

- Case 2.1: $x$ and $z$ are related by a single fuzzy set, i.e., there exists $B \in \mathbf{A}$ such that $x, z \in supp(B)$ and $\forall A_i \in \mathbf{A}$, $A_i \neq B$, $\min(A_i(x), A_i(z)) = 0$.

  If $y \notin supp(B)$, since $y \in supp_x \cup supp_z$, it either unrelated to $x$ or to $z$ (but not both) and thus, using Corollary 1.1, either $f^+(x, y) = 1$ or $f^+(y, z) = 1$ and therefore $f^+(x, z) \leq f^+(x, y) + f^+(y, z)$.

  Let us now suppose that $y \in supp(B)$. Without lack of generality, let us assume that $x \leq z$. From Corollary 1.2, if $x$ and $z$ belong to the core of $B$, we trivially have $f^+(x, z) = 0 \leq f^+(x, y) + f^+(y, z)$. Otherwise, we have $f^+(x, z) = \max(1 - B(x), 1 - B(z))$.

  Let us consider the latter case. There are two possibilities for $y$, it is either i) related to both $x$ and $z$ or ii) related to either $x$ or $z$ but not both.

  In i), $x, y, z \in supp(B)$ and thus $f^+(x, z) = \max(1 - B(x), 1 - B(z))$ $\leq \max(1-B(x), 1-B(y))+\max(1-B(y), 1-B(z)) = f^+(x, y)+f^+(y, z)$.

  In ii) either $y \in supp(B^-)$ or $y \in supp(B^+)$, where $B^-$ and $B^+$ are fuzzy sets in $\mathbf{A}$ to the left and right of $B$ respectively. If $y \in supp(B^-)$, $x$ and $y$ are related by two fuzzy sets, $B^-$ and $B$, and by Corollary 1.3 we have $f^+(x, y) =\mid B(x) - B(y) \mid$. Moreover, $y$ and $z$ are unrelated and thus $f^+(y, z) = 1$. Therefore, $f^+(x, z) = \max(1 - B(x), 1 - B(z)) \leq\mid B(x) - B(y) \mid + 1 = f^+(x, y) + f^+(y, z)$. The case in which $y \in supp(B^+)$ is proved in a similar manner.
- Case 2.2: $x$ and $z$ are related by exactly 2 fuzzy sets, i.e. when there exists two contiguous fuzzy sets $B_1, B_2$ in $\mathbf{A}$ such that $x, z \in supp(B_1) \cap supp(B_2)$ and for all $A_i \in \mathbf{A}$ such that $A_i \neq B_1$ and $A_i \neq B_2$, $\min(A_i(x), A_i(z)) = 0$.

  From Corollary 1.3, we have $f^+(x, z) =\mid B_1(x) - B_1(z) \mid$. Without lack of generality, let us suppose that $B_1$ stands to the left of $B_2$ and that $x \leq z$. We then have $B_1(x) \geq B_1(z)$ and thus $f^+(x, z) = B_1(x) - B_1(z)$.

  We verify two possibilities for $y$:
  - i) $y \in supp(B_1) \cap supp(B_2)$:
    We have to consider three orderings in what regards $y$ :
    - $y \leq x \leq z$: In this case, $B_1(y) \geq B_1(x) \geq B_1(z)$ and using Corollary 1.3, $f^+(x, y) = B_1(y) - B_1(x)$ and $f^+(y, z) = B_1(y) - B_1(z)$. But $f^+(x, z) = B_1(x) - B_1(z) \leq B_1(y) - B_1(x) + B_1(y) - B_1(z) = f^+(x, y) + f^+(y, z)$.
    - $x \leq y \leq z$: We have $B_1(x) \geq B_1(y) \geq B_1(z)$ and in a similar manner we obtain $f^+(x, z) = B_1(x) - B_1(z) = B_1(x) - B_1(y) + B_1(y) - B_1(z) = f^+(x, y) + f^+(y, z)$.
    - $x \leq z \leq y$: We have $B_1(x) \geq B_1(z) \geq B_1(y)$ and in a similar manner to $y \leq x \leq z$ we obtain $f^+(x, z) \leq f^+(x, y) + f^+(y, z)$.
  - ii) $y \notin supp(B_1) \cap supp(B_2)$:
    Since $y \in supp(B)$, either a) $y \in supp(B_1) - supp(B_2)$ or b) $y \in supp(B_2) - supp(B_1)$.

In a), $x$ and $y$ are related by two fuzzy sets, $B_1$ itself, and a fuzzy set to the left of $B_1$, which we will call $B_1^-$. Applying Corollary 1.3 we obtain $f^+(x,y) =\mid B_1(x) - B_1(y) \mid$. Moreover, since $y$ is unrelated to $z$, we obtain $f^+(y,z) = 1$ (see Corollary 1.1). Therefore $f^+(x,z) =\mid B_1(x) - B_1(z) \mid\leq\mid B_1(x) - B_1(y) \mid + 1 = f^+(x,y) + f^+(y,z)$. We obtain a similar result for the case b), which ends the proof.

**Corollary 3.** *Let $(\Omega, \preceq)$ be a total order,* **A** *a 2-Ruspini CFP$_\preceq$, and $f^+$ derived from* **A**. *If all fuzzy sets in* **A** *are triangular, $f^+$ satisfies the identity of indiscernibles property.*

*Proof.* To satisfy the identity of indiscernibles property, we should have $f^+(x,y) = 0$ if and only if $x = y$. Let us suppose that there exists two elements $p$ and $q$ in $\Omega$ such $f^+(p,q) = 0$ and $p \neq q$.

But $f^+(p,q) = 0$ only when two conditions are satisfied:

1. $\exists i, \min(A_i(p), A_i(q)) > 0$ and
2. $\sup_i \mid A_i(p) - A_i(q) \mid = 0$.

As seen in Corollary 1, any two elements in the domain may be either unrelated or related by 1 or 2 fuzzy sets. But $p$ and $q$ are necessarily related, due to the first condition above. Therefore, we have to analyse two cases:

- Case 1: $p$ and $q$ are related by a single fuzzy set $B$.
  In this case, as seen in Corollary 1.2, we have to verify weather both $p$ and $q$ belong to $core(B)$. But in a triangular fuzzy set $A \in \mathbf{A}$, $core(A) = c_A \in \Omega$, i.e. the core is a point. Since $p$ and $q$ are distinct and the core of $B$ is a point, $p$ and $q$ cannot not both belong to the $core(B)$. The remaining possibility is that $f^+(p,q) = \max(1-B(p), 1-B(q))$. But this expression go to 0 only when $B(p) = B(q) = 1$, which would mean that they both belong to the core of $B$, what contradicts the previous result.
- Case 2: $p$ and $q$ are related by two contiguous fuzzy sets $B_1$ and $B_2$.
  As seen in Corollary 1.3, we have $f^+(p,q) =\mid B_1(p) - B_1(q) \mid$. But in this case, $f^+(p,q) = 0$ only when $B_1(p) = B_1(q)$. Since $B_1$ is triangular, $B_1(p) = B_1(q)$ only when $p = q = core(B_1)$, which contradicts the hypothesis.

Since in both cases, we obtain a contradiction of the hypothesis, we conclude that the identity of indiscernibles property holds.

Function $f^+$ derived from a 2-Ruspini CFP$_\preceq$ **A** is not a distance in the general case, because identity of indiscernibles does not hold when there exists a fuzzy set $A$ in **A** such that $core(A) = [a,b]$, with $a, b \in \Omega$ and $a \neq b$, i.e, $A$ is either a trapezoidal fuzzy set or crisp interval. Indeed, in such a case, $\forall p, q \in [a,b]$ we have $f^+(p,q) = 0$. We now prove that $f^+$ is a pseudometric for any 2-Ruspini CFP$_\preceq$ **A** and a distance when all fuzzy sets in **A** are triangular.

**Proposition 1.** *Let $f^+$ be derived from a 2-Ruspini CFP$_\preceq$* **A** *on $\Omega$.*

1. *Function $f^+$ is a pseudometric.*
2. *Function $f^+$ is a distance when all fuzzy sets in* **A** *are triangular.*

*Proof.* It is straightforward to verify that $f^+$ satisfies symmetry, anti-reflexitivity and non-negativity, because $S^+$ is a symmetric and reflexive fuzzy relation. Moreover, in Corollary 2 we proved that the triangle inequality property holds for function $f^+$. This proves part 1.

We proved in Corollary 3 that $f^+$ satisfies the identity of indiscernibles property, when all fuzzy sets in $\mathbf{A}$ are triangular. This proves part 2, which completes the proof.

We now propose to use the arithmetic means to in order to extend $f^+$ to multi-dimensional domains. In Proposition 2, we prove that the resulting function has the same properties as in the one-dimensional domain.

**Proposition 2.** *Let $O = \Omega_1 \times ... \times \Omega_m$, where $\forall i, (\Omega_i, \preceq)$ is a total order. Let $\mathbf{A}_i$ be a 2-Ruspini CFP$_\preceq$ on $\Omega_i$ and $f_i^+$ be derived from $\mathbf{A}_i$. Let $f^+_{(\mu)} : O \to [0,1]$ be the extension of function $f^+$ to multidimensional domains, defined as*

$$f^+_{(\mu)}(x,y) = \mu(f_1^+(x_1,y_1), ..., f_m^+(x_m,y_m)),$$

*where $\mu : [0,1]^m \to [0,1]$ is the arithmetic means, i.e., $\mu(a_1, ..., a_m) = \frac{\sum_{1 \le i \le m} a_i}{m}$.*

1. *Function $f^+_{(\mu)}$ is a pseudometric.*
2. *Function $f^+_{(\mu)}$ is a distance when all fuzzy sets in $\mathbf{A}$ are triangular.*

*Proof.* Function $f^+_{(\mu)}$ trivially satisfies symmetry, anti-reflexitivity and non-negativity. We now prove that the triangle inequality property holds for $f^+_{(\mu)}$.

From the definitions of $f^+$ and $f^+_{(\mu)}$ and from Proposition 1, we have $f^+_{(\mu)}(x,z) = \frac{\sum_i f_i^+(x_i,z_i)}{m} \le \frac{\sum_i f_i^+(x_i,y_i)+f_i^+(y_i,z_i)}{m} = \frac{\sum_i f_i^+(x_i,y_i)}{m} + \frac{\sum_i f_i^+(y_i,z_i)}{m} = f^+_{(\mu)}(x,y) + f^+_{(\mu)}(y,z)$. Therefore, $f^+_{(\mu)}$ satisfies the triangle inequality and is thus a pseudometric. This proves part 1.

Let $x$ and $y$ be two elements in the m-dimensional space $O$. But from Proposition 1, when all fuzzy sets in each 2-Ruspini CFP$_\preceq$ $\mathbf{A}_i$, defined on its corresponding $\Omega_i$, is composed solely of triangular fuzzy sets, we have $\forall i(f_i^+(x_i,y_i) = 0 \Leftrightarrow x_i = y_i)$. But $\forall i(f_i^+(x_i,y_i) = 0 \Leftrightarrow x_i = y_i) \Rightarrow [\forall i(f_i^+(x_i,y_i) = 0) \Leftrightarrow \forall i(x_i = y_i)]$. Since we trivially have i) $x = y$ iff $\forall i \in [1,m], x_i = y_i$ and ii) $f_\mu^+(x,y) = 0$ iff $\forall i \in [1,m], f_i^+(x_i,y_i) = 0$, we thus have $(f_\mu^+(x,y) = 0 \Leftrightarrow x = y)$. Therefore $f^+_{(\mu)}$ satisfies the identity of indiscernibles and is thus a metric, when the $\mathbf{A}_i$'s are composed of triangular fuzzy sets, which concludes the proof.

## 4 Use of $f^+$ in a Classification Application

In the following, we briefly describe an experiment that illustrates the use of function $f^+$ in a land use classification task in the Brazilian Amazon region. The area of interest covers approximately 411 km$^2$ and in the municipality of Belterra, state of Pará, in the Brazilian Amazon region, partially contained in the National Forest of Tapajós. An intense occupation process occurred in the region along the BR-163 highway (Cuiabá-Santarém), with opening of roads to establish small farms, after deforestation of primary

forest areas [2]. As a result, there are mosaics of secondary vegetation in various stages, with pastures and cultivated areas embedded in a forest matrix [3].

In this application [5], 14 attributes have been considered, derived from either radar or optical satellite images, with 6 classes: forest, initial or intermediate regeneration, advanced regeneration or degraded forest, cultivated area, exposed soil, and pasture. The samples consist of 428 ground information based hand-made polygons. The attribute value for each polygon is the average of the values for the pixels composing it. The experiments have been done using 10 folders (9 for training and 1 for testing), partitioned as 2 sets containing 42 elements and 8 sets containing 43 elements.

Figure 1 brings the accuracy results for this application, considering k-NN with 1 to 19 neighbors, using the Euclidean distance (kNN_dE) and the Mahalanobis distance (kNN_dM), as well as function $f^{+}_{(\mu)}$ obtained from two types of partitions using 3 fuzzy sets each, a triangular one (kNN_dFtg) and a trapezoidal one (kNN_dFtz). In order to calculate $f^{+}$ for each attribute, the corresponding domain was reduced to the interval of the minimal and maximal sample values; both of which extended by 20%. For both the triangular and trapezoidal partitions, the fuzzy sets were homogeneously distributed on the reduced domain.

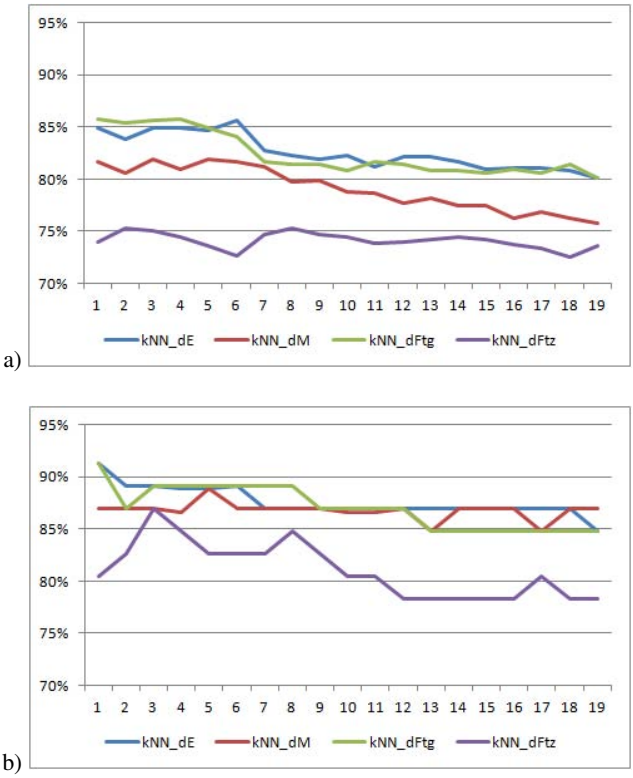



**Fig. 1.** Classification accuracy results for: a) k-NN average and b) k-NN maximum

We see from the figures that all methods had high accuracy and that the best results were obtained with the Euclidean distance and the use of $f_\mu^+$ for the chosen triangular partitions. We also see that the use of triangular partition with $f_\mu^+$, with the advantage of producing its best results with a small number of neighbours, contrarily to all of the other methods.

## 5 Conclusions

In this work, we have proposed the use of a pseudometric based on a particular type of fuzzy partition, that becomes a distance when the underlying partition is formed solely by triangular fuzzy sets. The extension of this function for multi-dimensional domains was also proposed. We have proved that this function derived from both trapezoidal and triangular partitions enjoy good properties, which make them formally suitable to be used in k-NN classifiers. Finally, we have shown a real-world experiment, in which this function obtained very good results, showing its practical use in applications. The obtained results are very promising and future work includes applying these functions in other experiments.

**Acknowledgements.** The authors acknowledge financial support from FAPESP and are very grateful to Mariane S. Reis for help on the Tapajós case study.

## References

1. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
2. Brazilian Institute of Environment and Renewable Natural Resources (IBAMA): Floresta Nacional do Tapajós Plano de Manejo, vol. I (2009) (in Portuguese), http://www.icmbio.gov.br/portal/images/stories/imgs-unidades-coservacao/flona_tapajoss.pdf (Date accessed: December 30, 2013)
3. Escada, M.I.S., Amaral, S., Rennó, C.D., Pinheiro, T.F.: Levantamento do uso e cobertura da terra e da rede de infraestrutura no distrito florestal da BR- 163. Repport INPE-15739-RPQ/824, INPE. S.J. Campos, Brazil (2009), http://urlib.net/sid.inpe.br/mtc-m18@80/2009/04.24.14.45 (Date accessed: December 30, 2013) (2009)
4. Korsrilabutr, T., Kijsirikul, B.: Pseudometrics for Nearest Neighbor Classification of Time Series Data. Engineering Journal 13 (May 2009), http://engj.org/index.php/ej/article/view/46 (Date accessed: December 30, 2013)
5. Martins-Bedê, F.T.: An extension to kNN classifier for multiple spaces, PhD Thesis. INPE. S.J. Campos, Brazil (in press, 2014) (in Portuguese)
6. Mendonça, J.H., Sandri, S., Martins-Bedê, F.T., Guimarães, R., Carvalho, O.: Training strategies for a fuzzy CBR cluster-based approach. Mathware & Soft Computing 20, 42–49 (2013)
7. Recasens, J.: Indistinguishability operators, modelling fuzzy equalities and fuzzy equivalence relations. STUDFUZZ, vol. 260. Springer (2011)
8. Sandri, S., Martins-Bedê, F.T.: A method for deriving order compatible fuzzy relations from convex fuzzy partitions. Fuzzy Sets and Systems (in press, 2014)
9. Theodoridis, S., Koutroumbas, K.: Pattern recognition, 3rd edn. Academic Press (2006)

# Quasi-Lovász Extensions on Bounded Chains

Miguel Couceiro and Jean-Luc Marichal

[1] LAMSADE - CNRS, Université Paris-Dauphine
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16, France
miguel.couceiro@dauphine.fr
[2] University of Luxembourg, Mathematics Research Unit
6, rue Richard Coudenhove-Kalergi
L–1359 Luxembourg, G.-D. Luxembourg
jean-luc.marichal@uni.lu

**Abstract.** We study quasi-Lovász extensions as mappings $f\colon C^n \to \mathbb{R}$ defined on a nonempty bounded chain $C$, and which can be factorized as $f(x_1,\ldots,x_n) = L(\varphi(x_1),\ldots,\varphi(x_n))$, where $L$ is the Lovász extension of a pseudo-Boolean function $\psi\colon \{0,1\}^n \to \mathbb{R}$ and $\varphi\colon C \to \mathbb{R}$ is an order-preserving function.

We axiomatize these mappings by natural extensions to properties considered in the authors' previous work. Our motivation is rooted in decision making under uncertainty: such quasi-Lovász extensions subsume overall preference functionals associated with discrete Choquet integrals whose variables take values on an ordinal scale $C$ and are transformed by a given utility function $\varphi\colon C \to \mathbb{R}$.

Furthermore, we make some remarks on possible lattice-based variants and bipolar extensions to be considered in an upcoming contribution by the authors.

**Keywords:** Aggregation function, discrete Choquet integral, Lovász extension, quasi-Lovász extension, comonotonic modularity.

**MSC Classes:** 39B22, 39B72 (Primary) 26B35 (Secondary).

## 1 Introduction and Preliminary Results

The discrete Choquet integral has been widely investigated in aggregation theory due to its many applications, for instance, in decision making (see the edited book [8]). A convenient way to introduce the discrete Choquet integral is via the concept of Lovász extension.

### 1.1 Lovász Extensions and the Discrete Choquet Integral

Let $C$ be a chain under a linear order $\leqslant$ or, equivalently, under a minimum $\wedge$ or a maximum $\vee$; for instance, $\mathbb{B} := \{0,1\}$ with the usual ordering given by

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 199–205, 2014.

$0 < 1$. Given a permutation $\sigma \in S_n$, where $S_n$ denotes the set of permutations on $[n] = \{1, \ldots, n\}$, we define

$$C^n_\sigma := \{\mathbf{x} = (x_1, \ldots, x_n) \in C^n : x_{\sigma(1)} \leqslant \cdots \leqslant x_{\sigma(n)}\}.$$

For every $A \subseteq [n]$, we denote by $\mathbf{1}_A$ the $n$-tuple whose $i$-th component is 1 if $i \in A$ and is 0 otherwise. For instance, $\mathbf{1}_\emptyset$ is the constant 0 tuple, denoted $\mathbf{0}$, whereas $\mathbf{1}_{[n]}$ is the constant 1 tuple.

Consider an $n$-place *pseudo-Boolean function*, i.e. a function $\psi\colon \mathbb{B}^n \to \mathbb{R}$, and define the set function $v_\psi\colon 2^{[n]} \to \mathbb{R}$ by $v_\psi(A) = \psi(\mathbf{1}_A)$ for every $A \subseteq [n]$. Hammer and Rudeanu [9] showed that such a function has a unique representation as a multilinear polynomial of $n$ variables

$$\psi(\mathbf{x}) = \sum_{A \subseteq [n]} a_\psi(A) \prod_{i \in A} x_i \,,$$

where the set function $a_\psi\colon 2^{[n]} \to \mathbb{R}$, called the *Möbius transform* of $v_\psi$, is defined by

$$a_\psi(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|}\, v_\psi(B).$$

The *Lovász extension* of a pseudo-Boolean function $\psi\colon \mathbb{B}^n \to \mathbb{R}$ is the function $L_\psi\colon \mathbb{R}^n \to \mathbb{R}$ whose restriction to each subdomain $\mathbb{R}^n_\sigma$ ($\sigma \in S_n$) is the unique affine function which agrees with $\psi$ at the $n+1$ vertices of the $n$-simplex $[0,1]^n \cap \mathbb{R}^n_\sigma$ (see [10,11]). We then have $L_\psi|_{\mathbb{B}^n} = \psi$.

It can be shown (see [7, §5.4.2]) that the Lovász extension of a pseudo-Boolean function $\psi\colon \mathbb{B}^n \to \mathbb{R}$ is the continuous function

$$L_\psi(\mathbf{x}) = \sum_{A \subseteq [n]} a_\psi(A) \bigwedge_{i \in A} x_i \,, \qquad \mathbf{x} \in \mathbb{R}^n. \tag{1}$$

Its restriction to $\mathbb{R}^n_\sigma$ is the affine function

$$L_\psi(\mathbf{x}) = \psi(\mathbf{0}) + \sum_{i \in [n]} x_{\sigma(i)} \big(v_\psi(A^\uparrow_\sigma(i)) - v_\psi(A^\uparrow_\sigma(i+1))\big), \qquad \mathbf{x} \in \mathbb{R}^n_\sigma, \tag{2}$$

or equivalently,

$$L_\psi(\mathbf{x}) = \psi(\mathbf{0}) + \sum_{i \in [n]} x_{\sigma(i)} \big(L_\psi(\mathbf{1}_{A^\uparrow_\sigma(i)}) - L_\psi(\mathbf{1}_{A^\uparrow_\sigma(i+1)})\big), \qquad \mathbf{x} \in \mathbb{R}^n_\sigma, \tag{3}$$

where $A^\uparrow_\sigma(i) = \{\sigma(i), \ldots, \sigma(n)\}$, with the convention that $A^\uparrow_\sigma(n+1) = \varnothing$. Indeed, for any $k \in [n+1]$, both sides of each of the equations (2) and (3) agree at $\mathbf{x} = \mathbf{1}_{A^\uparrow_\sigma(k)}$.

It is noteworthy that $L_\psi$ can also be represented by

$$L_\psi(\mathbf{x}) = \psi(\mathbf{0}) + \sum_{i \in [n]} x_{\sigma(i)} \big(L_\psi(-\mathbf{1}_{A^\downarrow_\sigma(i-1)}) - L_\psi(-\mathbf{1}_{A^\downarrow_\sigma(i)})\big), \qquad \mathbf{x} \in \mathbb{R}^n_\sigma,$$

where $A_\sigma^\downarrow(i) = \{\sigma(1), \ldots, \sigma(i)\}$, with the convention that $A_\sigma^\downarrow(0) = \varnothing$. Indeed, for any $k \in [n+1]$, by (3) we have

$$L_\psi(-\mathbf{1}_{A_\sigma^\downarrow(k-1)}) = \psi(\mathbf{0}) + L_\psi(\mathbf{1}_{A_\sigma^\uparrow(k)}) - L_\psi(\mathbf{1}_{A_\sigma^\uparrow(1)}).$$

A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is said to be a *Lovász extension* if there is a pseudo-Boolean function $\psi\colon \mathbb{B}^n \to \mathbb{R}$ such that $f = L_\psi$.

An $n$-place *Choquet integral* is a nondecreasing Lovász extension $L_\psi\colon \mathbb{R}^n \to \mathbb{R}$ such that $L_\psi(\mathbf{0}) = 0$. It is easy to see that a Lovász extension $L\colon \mathbb{R}^n \to \mathbb{R}$ is an $n$-place Choquet integral if and only if its underlying pseudo-Boolean function $\psi = L|_{\mathbb{B}^n}$ is nondecreasing and vanishes at the origin (see [7, §5.4]).

### 1.2 Quasi-Lovász Extensions on Bounded Chains

A generalization of the notion of Lovász extension of a pseudo-Boolean function was introduced in [4] and called "quasi-Lovász extension". In this paper we extend this notion to the following concept.

Let $C$ be a bounded chain under the usual operations $\wedge$ and $\vee$, and with least and greatest elements 0 and 1, respectively. We make no notational distinction between $0, 1 \in C$ and $0, 1 \in \mathbb{R}$; this notational abuse will not give raise to ambiguities.

We say that a function $f\colon C^n \to \mathbb{R}$ is a *quasi-Lovász extension* if there is a Lovász extension $L\colon \mathbb{R}^n \to \mathbb{R}$ and an order-preserving mapping $\varphi\colon C \to \mathbb{R}$ such that $f$ can be factorized into the composition

$$f(x_1, \ldots, x_n) = L(\varphi(x_1), \ldots, \varphi(x_n)). \tag{4}$$

For such a $\varphi\colon C \to \mathbb{R}$ we set $\varphi(C) := \{\varphi(x) : x \in C\}$ that is contained in some real interval $I$. To simplify our exposition we will assume that $\varphi(0) = 0$ and that $\varphi(C) \subseteq I = [0, a]$, for $a = \varphi(1)$ (shift by $\varphi(0)$ if necessary). Also, for every $f\colon C^n \to \mathbb{R}$, we denote by $f_0$ the function $f_0(\mathbf{x}) := f(\mathbf{x}) - f(\mathbf{0})$.

Such an aggregation function is used in decision under uncertainty, where $\varphi$ is a utility function and $f$ an overall preference functional. It is also used in multi-criteria decision making where the criteria are commensurate (i.e., expressed in a common scale). For a recent reference, see Bouyssou et al. [1].

## 2 Axiomatizations and Representations of Quasi-Lovász Extensions on Bounded Chains

We now provide an axiomatization of quasi-Lovász extensions on bounded chains in terms of comonotonic modularity and an ordinal variant of homogeneity, as well as provide a complete description of all possible factorizations (into a composition of a Lovász extension with an order-preserving unary mapping) when such a factorization exists.

### 2.1 Comonotonically Modular and Separable Functions

Recall that a function $f\colon C^n \to \mathbb{R}$ is said to be *modular* (or a *valuation*) if

$$f(\mathbf{x}) + f(\mathbf{x}') = f(\mathbf{x} \wedge \mathbf{x}') + f(\mathbf{x} \vee \mathbf{x}') \tag{5}$$

for every $\mathbf{x}, \mathbf{x}' \in C^n$, and where $\mathbf{x} \wedge \mathbf{x}'$ (resp. $\mathbf{x} \vee \mathbf{x}'$) denotes the $n$-tuple whose $i$th component is $x_i \wedge x_i' = \min(x_i, x_i')$ (resp. $x_i \vee x_i' = \max(x_i, x_i')$). It was proved (see Topkis [12, Thm 3.3]) that a function $f\colon C^n \to \mathbb{R}$ is modular if and only if it is *separable*, that is, there exist $n$ functions $f_i\colon C \to \mathbb{R}$, $i \in [n]$, such that $f = \sum_{i\in[n]} f_i$.[1] In particular, any 1-place function $f\colon C \to \mathbb{R}$ is modular.

We say that a function $f\colon C^n \to \mathbb{R}$ is *comonotonically modular* if (5) holds for every $\sigma \in S_n$ and $\mathbf{x}, \mathbf{x}' \in C^n_\sigma$.

**Fact 1.** *A function $f$ is comonotonically modular if and only if so is the function $f_0 = f - f(\mathbf{0})$.*

**Theorem 2.** *A function $f\colon C^n \to \mathbb{R}$ is comonotonically modular if and only if one or, equivalently, both of the conditions hold:*

*(i) there exists a function $g\colon C^n \to \mathbb{R}$ such that, for every $\sigma \in S_n$ and every $\mathbf{x} \in C^n_\sigma$,*

$$f(\mathbf{x}) = g(\mathbf{0}) + \sum_{i\in[n]} \big(g(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i)}) - g(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i+1)})\big). \tag{6}$$

*In this case, we can choose $g = f$.*

*(ii) $f$ is comonotonically separable, that is, for every $\sigma \in S_n$, there exist functions $f_i^\sigma : C \to \mathbb{R}$, $i \in [n]$, such that*

$$f(\mathbf{x}) = \sum_{i=1}^{n} f_i^\sigma(x_{\sigma(i)}) = \sum_{i=1}^{n} f^\sigma_{\sigma^{-1}(i)}(x_i), \qquad \mathbf{x} \in C^n_\sigma.$$

*In this case, we can choose $f_n^\sigma(x_{\sigma(n)}) = f(x_{\sigma(n)} \wedge \mathbf{1}_{\{\sigma(n)\}})$, and $f_i^\sigma(x_{\sigma(i)}) = f(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i)}) - f(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i+1)})$, for $i \in [n-1]$.*

*Proof.* It is easy to verify that $(i) \Rightarrow (ii)$. Moreover, each $f^\sigma_{\sigma^{-1}(i)}$ is clearly modular and hence comonotonically modular. Since the class of comonotonically modular functions is closed under addition, we have that $(ii)$ is sufficient. Thus, to complete the proof, it is enough to show that every comonotonically modular function $f$ satisfies $(i)$.

Let $\sigma \in S_n$ and $\mathbf{x} \in C^n_\sigma$. By comonotonic modularity, for every $i \in [n-1]$ we have

$$f(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i)}) + f(\mathbf{x}^0_{A^{\downarrow}_{\sigma}(i)}) = f(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i+1)}) + f(\mathbf{x}^0_{A^{\downarrow}_{\sigma}(i-1)}),$$

that is,

$$f(\mathbf{x}^0_{A^{\downarrow}_{\sigma}(i-1)}) = \big(f(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i)}) - f(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i+1)})\big) + f(\mathbf{x}^0_{A^{\downarrow}_{\sigma}(i)}). \tag{7}$$

By using (7) for $i = 1, \ldots, n-1$, we obtain (6) with $g = f$. □

[1] This result still holds in the more general framework where $f$ is defined on a product of chains.

### 2.2 Axiomatizations of Quasi-Lovász Extensions

To axiomatize the class of quasi-Lovász extensions, we will also make use of the following generalization of homogeneity. We say that $f\colon C^n \to \mathbb{R}$ is *weakly homogeneous* if there exists an order-preserving function $\varphi\colon C \to \mathbb{R}$ satisfying $\varphi(0) = 0$ such that $f(x \wedge \mathbf{1}_A) = \varphi(x) f(\mathbf{1}_A)$ for every $x \in C$ and every $A \subseteq [n]$. Note that every weakly homogeneous function $f$ satisfies $f(\mathbf{0}) = 0$.

The following two results are variants of Lemma 2 and Proposition 3 in [4]. and their proofs follow in complete analogy.

**Lemma 3.** *For every quasi-Lovász extension $f\colon C^n \to \mathbb{R}$, $f = L \circ \varphi$, we have*

$$f_0(x \wedge \mathbf{1}_A) = \varphi(x) L_0(\mathbf{1}_A), \qquad x \in C,\ A \subseteq [n]. \tag{8}$$

**Proposition 4.** *Let $f\colon C^n \to \mathbb{R}$ be a nonconstant quasi-Lovász extension, $f = L \circ \varphi$. Then the following conditions are equivalent.*

(*i*) *$f_0$ is weakly homogeneous.*
(*ii*) *There exists $A \subseteq [n]$ such that $f_0(\mathbf{1}_A) \neq 0$.*
(*iii*) *$\varphi(1) \neq 0$.*

*In this case we have $f_0(x \wedge \mathbf{1}_A) = \frac{\varphi(x)}{\varphi(1)} f_0(\mathbf{1}_A)$ for every $x \in C$ and every $A \subseteq [n]$.*

We can now provide axiomatizations of the class of quasi-Lovász extensions defined on bounded chains. The proof follows the same steps as in the proof of Theorem 14 in [4].

**Theorem 5.** *Let $f\colon C^n \to \mathbb{R}$ be a nonconstant function. Then, the following assertions are equivalent.*

(*i*) *$f$ is a quasi-Lovász extension and there exists $A \subseteq [n]$ such that $f_0(\mathbf{1}_A) \neq 0$.*
(*ii*) *$f$ is comonotonically modular and $f_0$ is weakly homogeneous.*
(*iii*) *There is an order-preserving function $\varphi_f\colon C \to \mathbb{R}$ satisfying $\varphi_f(0) = 0$ and $\varphi_f(1) = 1$ such that $f = L_{f|_{\{0,1\}^n}} \circ \varphi_f$.*

*Proof.* Let us prove that (*i*) $\Rightarrow$ (*ii*). By definition, we have $f = L \circ \varphi$, where $L\colon \mathbb{R}^n \to \mathbb{R}$ is a Lovász extension and $\varphi\colon C \to \mathbb{R}$ is an order-preserving function satisfying $\varphi(0) = 0$. By Proposition 4, $f_0$ is weakly homogeneous. Moreover, by (3) and (8) we have that, for every $\sigma \in S_n$ and every $\mathbf{x} \in C^n_\sigma$,

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{0}) + \sum_{i \in [n]} \varphi(x_{\sigma(i)}) \left( L_0(\mathbf{1}_{A^{\uparrow}_{\sigma}(i)}) - L_0(\mathbf{1}_{A^{\uparrow}_{\sigma}(i+1)}) \right) \\ &= f(\mathbf{0}) + \sum_{i \in [n]} \left( f(x_{\sigma(i)} \mathbf{1}_{A^{\uparrow}_{\sigma}(i)}) - f(x_{\sigma(i)} \mathbf{1}_{A^{\uparrow}_{\sigma}(i+1)}) \right). \end{aligned}$$

Theorem 2 then shows that $f$ is comonotonically modular.

Let us prove that $(ii) \Rightarrow (iii)$. Since $f$ is comonotonically modular, by Theorem 2 it follows that, for every $\sigma \in S_n$ and every $\mathbf{x} \in C^n_\sigma$,

$$f(\mathbf{x}) = f(\mathbf{0}) + \sum_{i\in[n]} \big(f(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i)}) - f(x_{\sigma(i)} \wedge \mathbf{1}_{A^{\uparrow}_{\sigma}(i+1)})\big),$$

and, since $f_0$ is weakly homogeneous,

$$f(\mathbf{x}) = f(\mathbf{0}) + \sum_{i\in[n]} \varphi_f(x_{\sigma(i)}) \big(f(\mathbf{1}_{A^{\uparrow}_{\sigma}(i)}) - f(\mathbf{1}_{A^{\uparrow}_{\sigma}(i+1)})\big) \tag{9}$$

for some order-preserving function $\varphi_f \colon C \to \mathbb{R}$ satisfying $\varphi_f(0) = 0$. By (3), we then obtain $f = L_{f|_{\{0,1\}^n}} \circ \varphi_f$. Finally, by (9) we have that, for every $A \subseteq [n]$,

$$f_0(\mathbf{1}_A) = \varphi_f(1) f_0(\mathbf{1}_A).$$

Since there exists $A \subseteq [n]$ such that $f_0(\mathbf{1}_A) \neq 0$ (for otherwise, we would have $f_0 \equiv 0$ by (9)), we obtain $\varphi_f(1) = 1$.

The implication $(iii) \Rightarrow (i)$ follows from Proposition 4. □

*Remark 6.* It is noteworthy that the class of quasi-polynomial functions on bounded chains [2] (or, more generally, on bounded distributive lattices [3]) is likewise axiomatizable in terms of comonotonic modularity by considering a lattice variant of weak homogeneity [6], namely: a function $f \colon C^n \to \mathbb{R}$ is said to be *weakly $\wedge$-homogeneous* if there exists an order-preserving function $\varphi \colon C \to \mathbb{R}$ such that $f(x \wedge \mathbf{1}_A) = \varphi(x) \wedge f(\mathbf{1}_A)$ for every $x \in C$ and every $A \subseteq [n]$.

### 2.3 Factorizations of Quasi-Lovász Extensions

We now describe all possible factorizations of $f$ into compositions of Lovász extensions with order-preserving functions.

**Theorem 7.** *Let $f \colon C^n \to \mathbb{R}$ be a quasi-Lovász extension, $f = L \circ \varphi$. Then there exists $A^* \subseteq [n]$ such that $f_0(\mathbf{1}_{A^*}) \neq 0$ if and only if there exists $a > 0$ such that $\varphi = a\,\varphi_f$ and $L_0 = \frac{1}{a}(L_{f|_{\mathbb{B}^n}})_0$.*

*Proof.* (Sufficiency) We have $f_0 = L_0 \circ \varphi = (L_{f|_{\{0,1\}^n}})_0 \circ \varphi_f$, and by Theorem 5 we see that the conditions are sufficient.

(Necessity) By Proposition 4, we have

$$\frac{\varphi(x)}{\varphi(1)} = \frac{f_0(x \wedge \mathbf{1}_{A^*})}{f_0(\mathbf{1}_{A^*})} = \varphi_f(x).$$

We then have $\varphi = a\,\varphi_f$ for some $a > 0$. Moreover, for every $\mathbf{x} \in \{0,1\}^n$, we have

$$\begin{aligned}(L_{f|_{\{0,1\}^n}})_0(\mathbf{x}) &= \big((L_{f|_{\{0,1\}^n}})_0 \circ \varphi_f\big)(\mathbf{x}) \;=\; f_0(\mathbf{x}) \;=\; (L_0 \circ \varphi)(\mathbf{x}) \\ &= a(L_0 \circ \varphi_f)(\mathbf{x}) \;=\; a\,L_0(\mathbf{x}).\end{aligned}$$

Since a Lovász extension is uniquely determined by its values on $\{0,1\}^n \subseteq \mathbb{R}$, we have $(L_{f|_{\{0,1\}^n}})_0 = a\,L_0$. □

## 3 Concluding Remarks and Directions for Further Research

We have axiomatized the class of quasi-Lovász extensions, considered in the wider setting of functions $f\colon\ C^n \to \mathbb{R}$ defined on a bounded chain $C$, thus partially generalizing the results presented in the proceedings of *IPMU2012* [5]. It remains now to consider the symmetric variant of the notion of quasi-Lovász extension defined on bipolar scales $C = C^{pos} \cup C^{neg}$ where $C^{pos}$ is a bounded chain and $C^{neg}$ is its "negative" copy.

**Acknowledgments.** This research is supported by the internal research project F1R-MTH-PUL-12RDO2 of the University of Luxembourg.

## References

1. Bouyssou, D., Dubois, D., Prade, P.M. (eds.): Decision-Making Process - Concepts and Methods. ISTE/John Wiley, London (2009)
2. Couceiro, M., Marichal, J.-L.: Axiomatizations of quasi-polynomial functions on bounded chains. Aeq. Math. 78(1-2), 195–213 (2009)
3. Couceiro, M., Marichal, J.-L.: Quasi-polynomial functions over bounded distributive lattices. Aeq. Math. 80(3), 319–334 (2010)
4. Couceiro, M., Marichal, J.-L.: Axiomatizations of quasi-Lovász extensions of pseudo-Boolean functions. Aeq. Math. 82, 213–231 (2011)
5. Couceiro, M., Marichal, J.-L.: Quasi-lovász extensions and their symmetric counterparts. In: Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R.R. (eds.) IPMU 2012, Part IV. CCIS, vol. 300, pp. 178–187. Springer, Heidelberg (2012)
6. Couceiro, M., Marichal, J.-L.: Discrete integrals based on comonotonic modularity. Axioms 2(3), 390–403 (2013)
7. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation functions. Encyclopedia of Mathematics and its Applications, vol. 127. Cambridge University Press, Cambridge (2009)
8. Grabisch, M., Murofushi, T., Sugeno, M. (eds.): Fuzzy measures and integrals - Theory and applications. STUDFUZZ, vol. 40. Physica-Verlag, Heidelberg (2000)
9. Hammer, P., Rudeanu, S.: Boolean methods in operations research and related areas. Springer, Heidelberg (1968)
10. Lovász, L.: Submodular functions and convexity. In: Mathematical Programming, 11th Int. Symp., Bonn, pp. 235–257 (1982)
11. Singer, I.: Extensions of functions of 0-1 variables and applications to combinatorial optimization. Numer. Funct. Anal. Optimization 7, 23–62 (1984)
12. Topkis, D.M.: Minimizing a submodular function on a lattice. Operations Research 26(2), 305–321 (1978)

# Efficient and Scalable Nonlinear Multiple Kernel Aggregation Using the Choquet Integral

Lequn Hu[1], Derek T. Anderson[1], Timothy C. Havens[2], and James M. Keller[3]

[1] Electrical and Computer Engineering, Mississippi State University, USA
[2] Electrical and Computer Engineering and Computer Science, Michigan Technological University, USA
[3] Electrical and Computer Engineering, University of Missouri, USA
lh432@msstate.edu, anderson@ece.msstate.edu, thavens@mtu.edu, kellerj@missouri.edu

**Abstract.** Previously, we investigated the definition and applicability of the *fuzzy integral* (FI) for nonlinear *multiple kernel* (MK) aggregation in pattern recognition. Kernel theory provides an elegant way to map multi-source heterogeneous data into a combined homogeneous (implicit) space in which aggregation can be carried out. The focus of our initial work was the Choquet FI, a per-matrix sorting based on the quality of a base learner and learning was restricted to the Sugeno $\lambda$-*fuzzy measure* (FM). Herein, we investigate what representations of FMs and FIs are valid and ideal for nonlinear MK aggregation. We also discuss the benefit of our approach over the linear convex sum MK formulation in machine learning. Furthermore, we study the Möbius transform and k-additive integral for scalable *MK learning* (MKL). Last, we discuss an extension to our *genetic algorithm* (GA) based MKL algorithm, called FIGA, with respect to a combination of multiple *light weight* FMs and FIs.

**Keywords:** Fuzzy integral, fuzzy measure, Möbius transform, multiple kernel learning, heterogeneous data fusion.

## 1 Introduction

Explosive growth in sensing and computing has given rise to numerous technological and mathematical dilemmas. Two well-known examples are *big data* and data diversity. Herein, we focus on the latter but in the context of a framework that can address the former as well. Consider the humanitarian effort of demining; the automatic identification and removal of hazards such as landmines and *improvised explosive devices* (IEDs) [1, 2]. These devices are responsible for injuring and claiming the lives of thousands of soldiers and civilians. The problem is that no single sensor or algorithm solves this challenging task. Instead, multiple sensors, multiple features, multiple algorithms and even human interaction/input is often critical for robust detection in different environments. However, information arising from multiple sensors, algorithms and people often result in great diversity, such as mixed-data type, multi-resolution (e.g., spatial and temporal),

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 206–215, 2014.

mixed-uncertainty, etc. An important question is, what is a well-grounded (non ad-hoc) way to carry out pattern analysis in light of such heterogeneity? This challenge is in no-way restricted to humanitarian demining. Another engineering example is combining multi-sensor, multi-band, multi-algorithm and even high-level human knowledge for wide area motion image analysis or earth observations using unmanned aerial vehicles. The point is, numerous challenges require the fusion and subsequent analysis of multi-source disparate data.

In prior work, we investigated the definition and applicability of the *fuzzy integral* (FI) for nonlinear *multiple kernel* (MK) aggregation in pattern recognition [3]. Kernel theory provides an elegant way to map multi-source heterogeneous data into a combined homogeneous (implicit) space where well-defined aggregation can be performed. The focus of our initial work was the Sugeno and Choquet FIs, a per-matrix sorting based on the quality of a base learner, e.g., a *support vector machine* (SVM), and learning was restricted to a Sugeno $\lambda$-*fuzzy measure* (FM). However, numerous questions remain: what types or representations of FIs are valid; are some representations better than others; what is an effective way to learn the FM for problems involving a large number of inputs; and what is the actual benefit of the FI for MK aggregation in comparison to other state-of-the-art MK techniques? Herein, we investigate these challenges.

This article is organized as follows. First, the FM, FI and the Möbius transform are reviewed. Next, MK aggregation and our prior *fuzzy integral MK learning* (FIMKL) work is described. The selection and representation of FI for nonlinear MK aggregation is then explored. Next, we investigate importance differences between FIMK and the popular machine learning *linear convex sum* (LCS) MKL formulation. The Möbius transform and the k-additive integral are then explored for efficient and scalable MKL. Last, we discuss the utilization of a combination of different *light weight* FMs in the context of our prior FIMKL *genetic algorithm* (GA) MKL algorithm, called FIGA.

## 2 Fuzzy Measure and Integral

The fusion of information using the Sugeno or Choquet FI has a rich history [4–8]. Depending on the problem domain, the input can be experts, sensors, features, similarities, pattern recognition algorithms, etc. The FI is defined with respect to the FM, a monotone (and often normal) measure. With respect to a set of $m$ information sources, $X = \{x_1, ..., x_m\}$, the FM encodes the (often subjective) *worth* of each subset in $2^X$.

**Definition 1 (Fuzzy Measure).** For a finite set of sources, $X$, the FM is a set-valued function $g : 2^X \to [0, 1]$ with the following conditions:

1. (Boundary condition) $g(\phi) = 0$,
2. (Monotonicity) If A, B $\subseteq X$ with A $\subseteq$ B, then $g(A) \leq g(B)$.

Note, if $X$ is an infinite set, there is a third condition guaranteeing continuity and we often assume $g(X) = 1$ as usual (although it is not necessary in general).

Numerous FI formulations have been proposed [4, 7, 9] for generalizability, differentiability and to address different types of uncertain data. Herein, we stick to the *conventional* (real-valued integrand and measure) Choquet integral.

**Definition 2 (Difference-in-Measure form of Choquet FI).** For a finite set of $m$ sources, FM $g$, and integrand $h : X \to \Re^+$, the discrete Choquet FI is

$$\int h \circ g = \sum_{i=1}^{m} \omega_i h\left(x_{\pi(i)}\right), \tag{1}$$

where $\omega_i = \left(G_{\pi(i)} - G_{\pi(i-1)}\right)$, $G_{\pi(i)} = g\left(\{x_{\pi(1)}, ..., x_{\pi(i)}\}\right)$, $G_{\pi(0)} = 0$, and $\pi(i)$ is a sorting on $X$ such that $h(x_{\pi(1)}) \geq \ldots \geq h(x_{\pi(m)})$.

Numerous measures exist, e.g., the Sugeno $\lambda$-FM, S-decomposable measures, possibility (and necessity) measures and k-additive FMs. The literature contains numerous ways to estimate their parameters from data, e.g., [10].

### 2.1 Möbius Transform

While the difference-in-measure and difference-in-integrand formulations are common, the FI can also be represented in terms of the Möbius transformation. The Möbius transform is of particular use for compactly expressing different formulas, e.g., Shapley and the k-additive measure/integral [11].

**Definition 3 (Möbius Transform).** The Möbius transformation of a FM $g$ is

$$\mathcal{M}(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} g(B), \qquad \forall A \subseteq X. \tag{2}$$

Note, the Möbius transformation is invertible via the Zeta transform,

$$g(A) = \sum_{B \subseteq A} \mathcal{M}(B), \qquad \forall A \subseteq X. \tag{3}$$

**Definition 4 (Möbius Transform Representation of the Choquet FI).** The Möbius transformation representation of the Choquet FI is

$$\int h \circ g = \sum_{A \subseteq X} \mathcal{M}(A) \bigwedge_{x_i \in A} h(x_i). \tag{4}$$

*Remark 1.* Equation (4) does not require sorting like Equation (1).

### 2.2 Multiple Kernel

In this section, we review basic concepts and definitions of kernels [12] needed for FIMK. First, assume that from each source of information in $X$, we measure a feature vector $\mathbf{x}$, where $\mathbf{x}_i$ describes the source $x_i$.[1]

**Definition 5 (Kernel).** Suppose we have a feature mapping $\phi : \mathcal{R}^d \to \mathcal{R}^{\mathcal{H}}$, where $d$ is the dimensionality of the input space and $\mathcal{R}^{\mathcal{H}}$ is a (higher-)dimensional space called the *Reproducing Kernel Hilbert Space* (RKHS). A kernel is the inner product function $\kappa : \mathcal{R}^d \times \mathcal{R}^d \to \mathcal{R}$, which represents the inner-product in $\mathcal{R}^{\mathcal{H}}$,

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle , \quad \forall \mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}. \tag{5}$$

**Definition 6 (Mercer Kernel).** Assume a kernel function $\kappa$ and finite data $\mathcal{X} = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$. The $n \times n$ matrix $K = [K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)]$, $i, j \in \{1, ..., n\}$, is a Gram matrix of inner products. Thus, $K$ is symmetric and *positive semi-definite* (PSD), $\mathbf{x}_i^T K \mathbf{x}_i \geq 0$, $\forall \mathbf{x}_i \in \mathcal{X}$. Since $K$ is PSD, all eigenvalues are non-negative.

This framework is particularly attractive for heterogeneous data fusion. Specifically, measuring similarity can be tailored to each individual domain. For example, if an input is a graph then a graph-based similarity kernel can be used. If another input is a histogram, then a kernel like the intersection kernel can be used. In effect, kernel theory provides a convenient way to map heterogeneous data into a homogeneous (implicit) space. In terms of aggregation, this is ideal. Aggregation can be engaged in the resultant homogeneous space to fuse multiple inputs. Furthermore, for homogeneous spaces, such as feature-valued data, kernel fusion is an elegant way to help mitigate challenges like feature vector size imbalance and feature vector scaling for the common method of combining feature-valued data (concatenation or *arraying*).

For readability, from now on we will denote features that are projected into the RKHS $\mathcal{R}^{\mathcal{H}}$ as $\phi_i$, where $\phi_i$ is equivalent to $\phi(\mathbf{x}_i)$. Furthermore, we will assume that each source in $X$ has either multiple features extracted from it or multiple kernels computed from a feature. We will use a super-script in this situation, viz., $\mathbf{x}_i^k$ is the feature vector describing source $x_i$ that is used to compute the $k$th kernel matrix $K_k$; thus, we use $\phi_i^k$ to denote $\phi(\mathbf{x}_i^k)$.

*Remark 2.* Let $K_1 = [\kappa_1(\mathbf{x}_i, \mathbf{x}_j)]$ and $K_2 = [\kappa_2(\mathbf{x}_i, \mathbf{x}_j)]$, $i, j \in \{1, ..., n\}$. The set of Mercer kernels is closed under the following (non-exhaustive) set of operations for $i, j \in \{1, ..., n\}$ [12]: $\mathcal{K}_{ij} = (K_1)_{ij} + (K_2)_{ij}$, $\mathcal{K}_{ij} = c(K_1)_{ij}, \forall c \geq 0$, $\mathcal{K}_{ij} = (K_1)_{ij} + c, \forall c \geq 0$, $\mathcal{K}_{ij} = (K_1)_{ij}(K_2)_{ij}$.

[1] Later, we will extend this idea to the scenario where one can have multiple feature vectors (really, multiple kernels) per source. This can manifest in one (or a combination) of two ways: i) different kernels can be computed from one feature type, or ii) a kernel can be computed from each of multiple heterogeneous feature types.

Kernel design takes a number of forms. In most scenarios, a single kernel is employed. Its parameters are generally specified manually or the selections of the kernel and its parameters are learned from data. However, the most widespread practice is to experimentally choose from a finite number of kernels and associated kernel parameters to pick a winner. In recent years, MK has been proposed to address the problem of finding a "best" kernel for a data set, and more recently, for focusing on the fusion and unique transformation of different sources (e.g., sensors and/or features). Specifically, MKL has emerged to find such kernel combinations automatically. Research has shown that the construction of a kernel from a number of base kernels allows for a more flexible encoding of domain knowledge from different sources or cues.

Linear aggregation is the strategy employed by most [13, 14]. It is based on a (potentially weighted) summation of base kernels, $\mathcal{K} = \sum_{k=1}^{m} \omega_k K_k$, where $\omega$ are weights, and $K_k = [\kappa_k(\mathbf{x}_i^k, \mathbf{x}_j^k)]$ is the kernel matrix produced by the $k$th feature extracted from the sources $X$. Many search for a single aggregate kernel to transform $x_i$ and $x_j$, thus one kernel function $\kappa$ is applied to all features extracted from $x_i$ and $x_j$. However, better performance may be obtained by placing each feature in its own space, viz., the $k$th feature vector gets its own kernel function $\kappa_k$. Approaches to MK aggregation differ in the way that restrictions are placed on the weights $\omega$. The most common categories include the linear sum ($\omega_k \in \mathcal{R}$), conic sum ($\omega_k \in \mathcal{R}^+$ ), and convex sum ($\omega_k \in \mathcal{R}^+$ and $\sum_{k=1}^{m} \omega_k = 1$). Compared to linear sum, conic and convex sums are appealing because they lead to weight (thus kernel and/or source) importance interpretation.

A few nonlinear aggregations methods have been proposed. However their representations and expressiveness are extremely limited and they are difficult to interpret. In [3], we investigated the Sugeno and Choquet FIs for MK. We proved that per-element aggregation is not theoretically valid. Semantically, per-element does make much sense, but mathematically—viz., in terms of production of a Mercer kernel—it is not valid (a counter proof was provided with respect to the maximum operator and negative eigenvalues). Instead, we proposed an alternative solution based on sorting at the *matrix level*. Assume each kernel matrix $K_k$ has a numeric "quality." As we showed, this can be computed, for example, by computing the classification accuracy of a base-learner that uses kernel $K_k$ (or by a learning algorithm such as a genetic algorithm). Let $\nu_k \in [0, 1]$ be the $k$th kernels *quality*. These qualities can be sorted, $\nu_{(1)} \geq \nu_{(2)} \geq ... \geq \nu_{(m)}$.

**Definition 7 (Difference-in-Measure Choquet FI for MK Aggregation).** Given $m$ base Mercer kernels, $\{\kappa_1, \ldots, \kappa_m\}$, FM $g$ and a sorting $\nu_{(1)} \geq \nu_{(2)} \geq ... \geq \nu_{(m)}$, the difference-in-measure Choquet FI

$$\mathcal{K}_{ij} = \sum_{k=1}^{m} (G_{\pi(k)} - G_{\pi(k-1)})(K_{\pi(k)})_{ij} = \sum_{k=1}^{m} \omega_k (K_{\pi(k)})_{ij}, \quad i, j \in \{1, ..., n\}, \quad (6)$$

produces a Mercer kernel as multiplication by positive scalar and addition are PSD preserving operations [3]. Since Equation (6) involves per-matrix sorting it can be compactly wrote in a simpler (linear algebra) form, $\mathcal{K} = \sum_{k=1}^{m} \omega_k K_{\pi(k)}$.

## 3 FIMK Insight and Scalable MKL

In this section, we address FIMKL expressibility and scalability. The first topic is how is FIMK different from the LCS MK formulation and why does it outperform other methods from machine learning? The second topic is a search for appropriate ways to represent FIMK to restrict its operation as the number of free parameters grow exponentially with respect to the number of inputs. This is significant in MKL as many explore the use of a relatively large number of kernels. A kernel can be applied to each sensor/source, feature index, or group (i.e., each bin in a histogram of gradients or the full descriptor). In addition, multiple kernels can be used for each of the above and different parameters are often explored. The point is, scalability is an important element of MKL.

### 3.1 Comparison of FIMK to Linear Convex Sum MK

The majority of machine learning MK research uses LCS form. It is often desired due to its advantage in optimization for MKL. One example is Bach's SMO-like algorithm for tackling convex and smooth minimization problems [15].

*Remark 3.* The weights in equations (1) and (6) are positive, $w_i \geq 0$, by definition (monotonicity) and their sum is 1, i.e.,

$$\sum_{k=1}^{m} w_k = \left(G_{\pi(m)} - G_{\pi(m-1)}\right) + ... + \left(G_{\pi(1)} - G_{\pi(0)}\right) = G_{\pi(m)} - G_{\pi(0)} = 1.$$

Both FIMK and LCS MK are type convex sum, i.e., $w_k \in \Re_+^m$ and $\sum_{k=1}^{m} w_k = 1$. However, one is linear, the other is not, and the weights are derived from the FM. The Choquet FI is capable of representing a much larger class of aggregation operators. For example, it is well known that the Choquet FI can produce, based on the selection of FM, the maximum, minimum, *ordered weighted average* (OWA), family of order statistics, etc. However, the machine learning LCS form is simply $m$ weights anchored to the individual inputs. The LCS is a subset (one of the aggregation operators) of the Choquet FI.

In [3], we reported improved SVM accuracies and lower standard deviations over the state-of-the-art, *MKL group lasso* (MKLGL) [14], on publically available benchmark data. We put forth a *genetic algorithm* (GA), called FIGA, based on learning the densities for the Sugeno $\lambda$-FM. An important question not addressed in our initial work is why exactly does FIMK perform notably better than LCS and, more specifically, MKLGL? Herein, we consider two possibilities. First is the expressibility of FIMK in terms of aggregation. Simply put, FIMK is nonlinear and more expressive, i.e., it can represent a much wider class of aggregation operators that can be specified or learned from data. Second is the learning algorithm, i.e., FIGA versus MKLGL. This is a more difficult topic to tackle mathematically. These two optimization algorithms operate in extremely different ways. Group lasso is an advanced approach designed to work on constrained problems, LCS type formulations. While it is relatively efficient and

mathematically elegant, it is simple to envision problems for which LCS is a inferior solution to a nonlinear or higher-order polynomial solution. On the other hand, GAs are a family of meta-heuristic optimization techniques that can operate on extremely challenging optimization surfaces. They exist to help avoid, potentially, pitfalls of early convergence to local minimia relative to a given initialization. Arguments can be made for both, i.e., efficiency versus expressibility. Globally, it is not likely this question has an answer (one approach *better* than the other). At the end of this article we show preliminary results for reducing FIGA to learning a simpler LCS solution to (empirically) understand what impact optimization had on our previous SVM-based experiments. However, first we investigate an additional compact way of representing the FI for MKL.

### 3.2 k-additive Choquet Integral for FIMK

In this section, we explore a definition of the Choquet FI for MK aggregation under the Möbius transform. If we attempt to directly analyze Equation (4) in terms of a valid Mercer kernel producing aggregation operator then we run into the same problem as before of per-element sorting and subsequently the existance of a FM of all ones ($g(A) = 1, \forall A \subseteq X$ such that $A \neq \phi$) that results in the maximum (known to not preserve PSD). Furthermore, the Möbius transform values can be positive, zero or negative. We know multiplication by positive scalars results in preservation of PSD, but we cannot guarantee PSD preservation in general for multiplication by negative numbers. Note, based on our proposed matrix-level sorting with respect to a base learner, this condition also arises for the common difference-in-integrand form of the Choquet FI and MK aggregation. However, we proved that the difference-in-measure form does indeed guarantee PSD preservation. Furthermore, we know that the difference-in-measure and difference-in-integrand are equivalent, they can algebraically be rewritten in terms of one another. We also know the Möbius transform form is equilivant to the difference-in-measure form of the Choquet FI. Two representations do not make it clear if the Choquet FI is valid with respect to per-matrix sorting. However, we proved (see [3]) one of these three forms and the other two are valid as well as they are simply re-formulations of one another.

Specifically, Equation (4) is not guaranteed to be valid because minimum is performed on a per-element basis (counter proof is trivial). We consider a slight reformulation of the Möbius transform, in combination with k-additivity, for MK aggregation that preserves the PSD property of matrices. Here, k-additivity is explored as it limits interaction between subsets to size $|A| \leq k, A \subseteq X$.

**Definition 8 (k-additive FI for MK).** The k-additive form of the Choquet FI for MK aggregation on a per-matrix basis (in terms of linear algebra) is

$$\mathcal{K} = \sum_{A \subseteq X, |A| \leq k} \mathcal{M}(A) \bigwedge_{x_i \in A} K_i, \tag{7}$$

where $\bigwedge$ is a per-matrix operator with respect to the associated $\nu_{(i)}$ values.

The main reason for considering the Möbius transform, outside of academic curiosity, is MKL. In MKL, it is often the case that we consider a large number of kernels. Different kernels for different sensors, features, feature groups, sets of parameters for kernels, etc. This presents a serious challenge for FIMK. Namely, FIMK is driven by the FM and the FM is exponential in the number of inputs. For example, [14] considered 793 kernels (different kernels for each individual feature and different kernels for the entire feature vector). While some might consider this rather extreme, such an approach would result in an unsolvable (intractable) problem if the full lattice was desired in FIMK, i.e., $2^{793} - 2$ free parameters. Constraints need be imposed to make FIMK scalable in this domain. For example, the 2-additive FI only requires $m + \binom{m}{2}$ parameters, where $\binom{m}{2}$ is the number of $2-$combinations. Regardless, this problem is still constrained by the monotonicity constraints and boundary conditions.

*Example 1 (Maximum Operator).* Consider a FM in which $g(A) = 1, \forall A \subseteq X$. Furthermore, let $m = 3$. One obtains $\mathcal{M}(\phi) = 0$, $\mathcal{M}(x_i) = 1$, $\mathcal{M}(\{x_i, x_j\}) = -1$ and $\mathcal{M}(X) = 1$. With respect to Equation (7), we get

$$\begin{aligned}\mathcal{K} &= \mathcal{M}(x_1)K_1 + ... + \mathcal{M}(\{x_1, x_2\})\,(K_1 \wedge K_2) + ... + (K_1 \wedge K_2 \wedge K_3) \\ &= K_1 + K_2 + K_3 - (K_1 \wedge K_2) - (K_1 \wedge K_3) - (K_2 \wedge K_3) + (K_1 \wedge K_2 \wedge K_3)\,.\end{aligned}$$

Furthermore, for a base learner sorting like $\nu_3 \geq \nu_2 \geq \nu_1$,

$$\mathcal{K} = K_1 + K_2 + K_3 - K_1 - K_1 - K_2 + K_1 = K_3,$$

i.e., we get the *maximum* (with respect to our base learners) of our kernels.

*Example 2 (Average).* Consider $m$ sources and a FM in which $g(\phi) = 0$, $g(X) = 1$, and $g(A) = \frac{|A|}{|X|}$ for $A \subset X \setminus \phi$. We obtain $\mathcal{M}(\phi) = 0$, $\mathcal{M}(x_i) = \frac{1}{|X|}$, $\mathcal{M}(A) = 0$ for $A \subseteq X, |A| > 1$, and thus $\mathcal{K} = \frac{1}{m}K_1 + ... + \frac{1}{m}K_m$.

## 4 Multi-measure FIGA and Preliminary Results

The primary purpose of this article is to explore different *compact* representations of the FM and FI for MKL. We experimentally investigate if there is any observable benefit for SVM-based classification in pattern classification. Our results are compared to the LCS form and MKLGL on public domain benchmark data. Specifically, we explore an extension to FIGA. Herein, we explore a combination of the k-additive integral, the possibility measure and an OWA for MKL. The Sugeno $\lambda$-FM is not used as $\lambda$ becomes difficult to accurately calculate in a computer for high order polynomials ($(m-1)^{th}$ order). These are three different *compact* measures (i.e., they involve relatively few number of free parameters versus $2^m - 2$) that together provide a wide range of possibilities for learning an effective MK aggregation strategy. In the case of the possibility measure and an OWA FM, $m$ values are learned (densities for the former and the weights in

the latter). To begin, a chromosome is assigned to one of these three *types*. We use standard one point crossover between the densities (weights in the case of the OWA). In the case of an OWA, the values are normalized so they sum to 1. When a k-additive FM is crossed with another type of FM, just the $m$ densities are crossed and monotonicity constraints are checked. Any violated constraints are "repaired" by simply assigning the smallest possible valid value. We also use elitism so that the best solution is kept every generation.

Below, FIGA is also used to search for the weights in an LCS to understand the result of MKLGL versus the proposed GA framework. Specifically, what contribution do we observe with respect to the GA instead of linear versus nonlinear aggregation? In the case of the GA, 20 kernels are used on the full feature vector: the dot product, the RBF (different multiples of 10 relative to $\frac{1}{d}$, where $d$ is the feature dimensionality) and the polynomial kernel (of second, third and fourth order). We also note that the MKLGL article used a much greater number of kernels, 117, 442 and 793 respectively.

**Table 1.** Comparison of FIGA to MKLGL and FIMK to LCS form

| | **Method** | **Breast** | **Ionosphere** | **Sonar** |
|---|---|---|---|---|
| Classification Accuracy $[0, 100]\%$ | MKLGL reported in [14] | $96.6 \pm 1.2$ | $92.0 \pm 2.9$ | $82.0 \pm 6.8$ |
| | FIGA: just LCS form | $97.95 \pm 0.17$ | $94.23 \pm 0.50$ | $91.03 \pm 2.70$ |
| | FIGA: combination of FMs/FIs | $98.00 \pm 0.14$ | $95.40 \pm 0.47$ | $92.86 \pm 1.17$ |

Table 1 tells the following story. First, it is clear that the GA approaches are more effective than the MKLGL approach, even though the GA approaches use fewer component kernels. Note that the FIGA approaches achieve a mean improvement of about 10% over MKLGL on the Sonar data set. The performance of FIGA comes at a cost though, as MKLGL is much faster in terms of actual running time than FIGA. Second, we see that FIGA using a combination of FM/FIs is somewhat more effective than the FIGA LCS form. These findings are not surprising as our intuition tells us that the nonlinear aggregation allowed by the FM/FI formulation is more flexible than just the LCS aggregation; hence, these results reinforce our expectation. Furthermore, FIGA using the combination of different compact FMs and FIs leads to improved performance at no real additional cost over the FIGA using just an LCS aggregation. Overall, these results are not surprising as different data sets require different solutions, and while an LCS may be sufficient for a given problem, it may not be appropriate for a different problem. Also, it should be noted that the FM/FI formulation includes LCS aggregation as a subset of its possible solutions; hence, when LCS is appropriate the FM/FI aggregation can mimic the LCS. In summary, these experiments suggest that the learner (GA vs GL) appears to be the most important improvement factor, followed by a slight improvement by using the nonlinear FM/FI aggregation versus LCS.

## 5 Conclusion and Future Work

In summary, we explored different compact FMs and FIs and their combination for MKL in the context of a GA for pattern recognition. We compared this framework to the LCS formulation and MKLGL optimization approach from machine learning. These contributions led to performance benefit in terms of SVM-based classification on benchmark data sets. In future work, we will explore additional, (constrained) efficient ways of learning the FM for FIMKL. We will also explore more efficient non-GA solvers relative to a set of FMs.

## References


1. Anderson, D.T., Stone, K., Keller, J.M., Spain, C.: Combination of Anomaly Algorithms and Image Features for Explosive Hazard Detection in Forward Looking Infrared Imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5(1), 313–323 (2012)
2. Havens, T.C., Keller, J.M., Stone, K., Anderson, D.T., Ho, K.C., Ton, T.T., Wong, D.C., Soumekh, M.: Multiple kernel learning for explosive hazards detection in forward-looking ground-penetrating radar. SPIE Defense, Security, and Sensing (2012), http://dx.doi.org/10.1117/12.920482
3. Hu, L., Anderson, D.T., Havens, T.C.: Fuzzy Integral for Multiple Kernel Aggregation. In: IEEE International Conference on Fuzzy Systems (2013)
4. Grabisch, M., Nguyen, E., Walker, E.: Fundamentals of uncertainty calculi with applications to fuzzy inference. Kluwer Academic, Dordrecht (1995)
5. Grabisch, M., Murofushi, T., Sugeno, M.: Fuzzy measures and integrals: theory and applications. STUDFUZZ. Physica-Verlag (2000)
6. Tahani, H., Keller, J.: Information fusion in computer vision using the fuzzy integral. IEEE Trans. on Systems, Man, and Cyber. 20, 733–741 (1990)
7. Anderson, D.T., Havens, T.C., Wagner, C., Keller, J.M., Anderson, M., Wescott, D.: Extension of the Fuzzy Integral for General Fuzzy Set-Valued Information. IEEE Trans. on Fuzzy Systems (2014), doi:10.1109/TFUZZ.2014.2302479
8. Grabisch, M., Labreuche, C.: A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. 4OR: A Quarterly Journal of Operations Research 6, 1–44 (2008)
9. Sugeno, M.: Theory of fuzzy integrals and its application, Ph.D. thesis, Tokyo Institute of Technology (1974)
10. Mendez-Vazquez, A., Gader, P.D.: Learning Fuzzy Measure Parameters by Logistic LASSO. In: Proceedings of the North American Fuzzy Information Processing Society Meeng, NAFIPS, New York, NY, pp. 1–7 (2008)
11. Grabisch, M.: k-order additive discrete fuzzy measures and their representation. Fuzzy Sets and Systems 92, 167–189 (1997)
12. Scholkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press (2001)
13. Gonen, M., Alpaydin, E.: Multiple kernel learning algorithms. Journal of Machine Learning Research 12, 2211–2268 (2011)
14. Xu, Z., Jin, R., Yang, H., King, I., Lyu, M.: Simple and Efficient Multiple Kernel Learning by Group Lasso. In: Int'l Conference on Machine Learning (2010)
15. Bach, F., Lanckriet, G., Jordan, M.: Multiple kernel learning conic duality, and the SMO algorithm. In: Int'l Conference on Machine Learning, pp. 41–48 (2004)

# On the Informational Comparison of Qualitative Fuzzy Measures

Didier Dubois[1], Henri Prade[1], and Agnès Rico[2]

[1] IRIT, CNRS and Université de Toulouse, France
{dubois,prade}@irit.fr
[2] ERIC, Université de Lyon, France
agnes.rico@univ-lyon1.fr

**Abstract.** Fuzzy measures or capacities are the most general representation of uncertainty functions. However, this general class has been little explored from the point of view of its information content, when degrees of uncertainty are not supposed to be numerical, and belong to a finite qualitative scale, except in the case of possibility or necessity measures. The thrust of the paper is to define an ordering relation on the set of qualitative capacities expressing the idea that one is more informative than another, in agreement with the possibilistic notion of relative specificity. To this aim, we show that the class of qualitative capacities can be partitioned into equivalence classes of functions containing the same amount of information. They only differ by the underlying epistemic attitude such as pessimism or optimism. A meaningful information ordering between capacities can be defined on the basis of the most pessimistic (resp. optimistic) representatives of their equivalence classes. It is shown that, while qualitative capacities bear strong similarities to belief functions, such an analogy can be misleading when it comes to information content.

**Keywords:** Fuzzy measures, possibility theory, qualitative reasoning, information content.

## 1 Introduction

Qualitative fuzzy measures (or q-capacities) are monotonic set-functions on a finite set with a range in a finite totally ordered set. They are instrumental in the representation of uncertainty of events, or yet, of the relative weights of groups of criteria in multicriteria evaluation, in the non-numerical environment, when likelihood or value scales are just totally or partially ordered sets, e.g. complete lattices [6]. An important issue to be clarified, if q-capacites are to be seriously considered as a tool for representing uncertainty, is the one of information content, and more generally the comparison of q-capacities in terms of their relative information content.

Important special cases of q-capacities are possibility and necessity measures [4]. For the latter set-functions, there is an abundant literature concerning information comparison, based on the notion of relative specificity [14,4]. Namely,

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 216–225, 2014.

a possibility (or a necessity) measure is entirely characterized by a possibility distribution over the elementary events, that defines a fuzzy set, and information comparison is carried out using fuzzy set inclusion, whereby a fuzzy set is more specific than another one, if and only if the former is a subset of the latter. However, there is almost no literature on the informational comparison of q-capacities in the general case. Some authors [10] have proposed notions of entropy for discrete capacities that evaluate the diversity of the coefficients appearing in the set-function. Our aim is rather to extend the specificity ordering to general q-capacities.

To do so, one may get inspiration from the theory of belief functions, where several informational orderings of various strength have been defined [13,3]. A belief function is a $\infty$-monotone set-function, that can be defined by means of a probability distribution over a power set (its Moebius transform), the degree of belief of an event summing the probability weights of all subevents that imply it. A belief function is then more informative than another if the former is eventwise greater than the latter. Another stronger definition, called specialisation, is based on the inclusion of focal sets. In the qualitative setting, it is also possible to compare q-capacities in an eventwise way. Besides, a qualitative counterpart of the Moebius transform can be defined, and a qualitative form of specialisation can be defined as well [12]. However, when indistinctly applied to q-capacites (like possibility and necessity measures), these definitions, even if formally well-defined, seem to baffle intuition.

In this paper, we try to provide a meaningful notion of information comparison between capacities, noticing that q-capacities differ not only by their information content, but also by the range of attitudes towards uncertainty they may encode. For instance, based on a given possibility distribution, the possibility measure is optimistic and the necessity measure is pessimistic [8], but one may argue they have the same information content.

## 2 Framework and Notations

We consider a finite set (of states, criteria, etc.) $S = \{s_1, \cdots, s_n\}$ and a finite totally ordered scale $L$ with top 1 and bottom 0. Let min denote the minimum, max the maximum. Moreover $L$ is equipped with an involutive negation $\nu$. A q-capacity is a mapping $\gamma : 2^S \to L$ such that $\gamma(\emptyset) = 0$, $\gamma(S) = 1$, and if $A \subseteq B$ then $\gamma(A) \leq \gamma(B)$. When $L = \{0, 1\}$, $\gamma$ is called a Boolean capacity.

A special case of capacity is a possibility measure defined with a possibility distribution $\pi : S \to L$. The possibility measure is defined by $\Pi(A) = \max_{s \in A} \pi(s)$. The value $\pi(s)$ is understood as the possibility that $s$ be the actual state of the world. Precise information corresponds to the situation where $\exists s^*, \pi(s^*) = 1$ and $\forall s \neq s^*, \pi(s) = 0$. Complete ignorance is represented by the vacuous possibility distribution $\pi^?$ such that $\forall s \in S$, $\pi^?(s) = 1$. A possibility distribution $\pi$ is more specific than another possibility distribution $\rho$ if $\forall s \in S, \pi(s) \leq \rho(s)$ [14,4]. This definition makes perfect sense since the set of possible values represented by $\pi$ is smaller, hence more precise, than the one represented by $\rho$.

The q-capacity *conjugate* of $\gamma$, denoted by $\gamma^c$, is defined using the involutive negation $\nu$ by $\gamma^c(A) = \nu(\gamma(A^c))\ \forall A \subseteq S$, where $A^c$ is the complement of the set $A$. In particular, the conjugate of a possibility measure is a necessity measure which is of the form $N(A) = \nu(\max_{s \notin A} \pi(s))$.

The *inner qualitative Moebius transform* of $\gamma$ is a mapping $\gamma_{\#} : 2^S \to L$ defined by $\gamma_{\#}(E) = \gamma(E)$ if $\gamma(E) > \max_{B \subset E} \gamma(B)$ and 0 otherwise [9,11]. It contains the minimal information that is sufficient to reconstruct $\gamma$ as $\gamma(A) = \max_{E \subseteq A} \gamma_{\#}(E)$. Let $\mathcal{F}^{\gamma} = \{E, \gamma_{\#}(E) > 0\}$ denote the family of so-called *focal sets* associated to $\gamma$. Note that the inner qualitative Moebius transform of $\gamma$ is strictly monotonic with inclusion on $\mathcal{F}^{\gamma}$. The inner qualitative Moebius transforms of a possibility measure $\Pi$ coincides with its possibility distribution $\pi$ (focal sets are singletons) while the focal sets of a necessity measure $N$ are nested (the cuts of $\pi$).

## 3 Can We Transfer Quantitative Definitions of Information Comparison to the Qualitative Setting?

When comparing capacities $\gamma$ and $\delta$, the inequality $\gamma \leq \delta$ is not always expressing something relevant about how informative $\gamma$ is with respect to $\delta$. Indeed, for instance if $\gamma$ is the vacuous possibility function $\Pi^?(A) = 1$ if $A \neq \emptyset$ and $\delta$ is the vacuous necessity function $N^?(A) = 0$ if $A \neq \Omega$, we have $\Pi^? > N^?$. However, they have exactly the same information content since based on the vacuous possibility distribution assigning 1 to all elements of $S$ and they are maximally uniformative among other possibility and necessity functions.

In the numerical setting, information comparison relations exist especially in the setting of belief functions. A belief function on $S$ is defined by means of a probability distribution $m$ over $2^S \setminus \{\emptyset\}$ as

$$Bel(A) = \sum_{E, E \subseteq A} m(E), \forall A \subseteq S. \tag{1}$$

The conjugate functions are plausibility measures defined by $Pl(A) = 1 - Bel(A^c)$. $Bel(A) = 1$ expresses full certainty of $A$, since then $Pl(A^c) = 0$ expresses that the complement $A^c$ is impossible.

There are several definitions of information comparison for belief functions. A belief function $Bel_1$ is said to be more informative (in the wide sense) than $Bel_2$ if $\forall A \subseteq S, Bel_1(A) \geq Bel_2(A)$. This is due to the fact that $Bel_1$ assigns greater degrees of certainty to events (while $Bel(A) = 0$ expresses no information). In terms of imprecise probabilities, it is equivalent to have the inclusion $\{P : P \geq Bel_1\} \subseteq \{P : P \geq Bel_2\}$. And notice that for plausibility functions the inequality is reversed ($Pl_1(A) \leq Pl_2(A)$).

Another information comparison method is based on *specialisation*, that relies on the mass assignments : $m_1$ is a specialization of $m_2$ (denoted by $m_1 \sqsubseteq_s m_2$) if and only if there exists a joint mass $x(A, B)$ with marginals $m_1$ and $m_2$, such that $x(A, B) = 0$ whenever $A \not\subseteq B, A \in \mathcal{F}_1, B \in \mathcal{F}_2$. It expresses inclusion between

focal sets of $Bel_1$ and $Bel_2$, that is sets $E$ with $m_i(E) > 0$. It can be checked that $m_1 \sqsubseteq_s m_2$ implies that $Bel_1$ is more informative (in the wide sense) than $Bel_2$, but not conversely.

There is a formal analogy between q-capacities and belief functions, since $\gamma(A) = \max_{E,E\subseteq A} \gamma_\#(E)$, whereby $\gamma_\#$ plays the role of a mass assignment, and $\sum$ turns into max. This is why it is then tempting to consider $\gamma$ as more informative than $\delta$ whenever $\forall A \subseteq S, \gamma(A) \geq \delta(A)$. This is all the more natural as the following result, adapted from [12], holds:

**Proposition 1.** *$\forall A \subseteq S, \gamma(A) \geq \delta(A)$ if and only if $\forall F \in \mathcal{F}^\delta, \exists E \in \mathcal{F}^\gamma$ s.t. $E \subseteq F$, $\gamma_\#(E) \geq \delta_\#(F)$.*

**Proof:** $\gamma(A) \geq \delta(A)$ writes $\max_{E,E\subseteq A} \gamma_\#(E) \geq \max_{F,F\subseteq A} \delta_\#(F)$. Suppose $A = F$ is a focal set of $\delta$. Then the latter is equivalent to $\max_{E,E\subseteq F} \gamma_\#(E) \geq \delta_\#(F)$, hence to $\forall F \in \mathcal{F}^\delta, \exists E \in \mathcal{F}^\gamma$ s.t. $E \subseteq F$ and $\gamma_\#(E) \geq \delta_\#(F)$.
Conversely, suppose the latter holds. If $\delta(A) = 0$ then the result is obvious. If $\delta(A) \neq 0$ hence let us consider $F$ the focal element included in $A$ such that $\delta(A) = \delta_\#(F)$. There exists $E$ a focal element of $\gamma$ included in $F$ such that $\gamma_\#(E) \geq \delta(A)$. We have $E \subseteq F \subseteq A$ so $\gamma(A) \geq \gamma_\#(E) \geq \delta(A)$. □

The condition $\forall F \in \mathcal{F}^\delta, \exists E \in \mathcal{F}^\gamma$ s.t. $E \subseteq F$, $\gamma_\#(E) \geq \delta_\#(F)$ is clearly a qualitative rendering of the specialisation relation. It formally means that for any focal set $F$ of $\delta$ there is a more important and more precise focal subset of $\gamma$, that explains the domination of $\gamma$ over $\delta$. Here, the two definitions of informational comparison are equivalent, which departs from the quantitative case.

However, this result is misleading. Neither the values of their lower Moebius transforms $\gamma_\#$ and $\delta_\#$, nor the size of focal sets $A, B$ with $\gamma_\#(A) > 0$ and $\delta_\#(B) > 0$ tell us much on their information content. For instance, focal sets of the vacuous possibility measure $\Pi^?$ are all singletons, and the unique focal set of $N^?$ is the whole set $S$. Viewing $\gamma$ as a counterpart to belief functions w.r.t. $\gamma_\#$ is thus not appropriate.

In such a context this article focuses on a basic question: When does a q-capacity represent the idea of certainty like belief functions, when does it represent plausibility (the conjugate of a belief functions)? In other words, when it is uncertainty-averse or pessimistic ? When it is uncertainty-prone or optimistic ?

## 4 Optimistic and Pessimistic q-Capacities

Given a possibility distribution $\pi$ the corresponding $\Pi$ is optimistic in the sense that $\Pi(\{s\}) = 1$ as soon as $s$ is fully possible (among other states); and its corresponding conjugate $N = \Pi^c$ is pessimistic, in particular $N(\{s\}) = 0, \forall s \in S$ as soon as two distinct elements in $S$ are fully possible. More generally, $N(A) \leq \Pi(A), \forall A \subseteq S$ and more specifically $N(A) > 0$ implies $\Pi(A) = 1$. Finally, $N(A) = 1$ expresses full certainty while $\Pi(A) = 1$ just expresses a lack of surprise for the occurrence of $A$. Likewise, belief functions are pessimistic while

their conjugate plausibility functions, which rely on the same mass function, are optimistic.

The above considerations motivate the following definition.

**Definition 1.** *A q-capacity $\gamma$ is said to be* pessimistic *(resp. optimistic) if $\gamma \leq \gamma^c$ (resp. if $\gamma \geq \gamma^c$).*

This definition first appears in [6] where a pessimistic (resp. optimistic) capacity is called uncertainty-averse (resp: uncertainty-prone). It is easy to see that:

- a q-capacity may be neither pessimistic nor optimistic. There may exist $A, B$ such that $\gamma(A) < \gamma^c(A)$ (pessimistic for $A$), and $\gamma(B) > \gamma^c(B)$ (optimistic for $B$).
- a q-capacity may be both : $\gamma = \gamma^c$ is possible. For instance with Boolean q-capacities ($L = \{0, 1\}$) on a space with $2n + 1$ elements, $\gamma_n(A) = 1$ if $|A| > n$ and 0 otherwise.

To check if a capacity is pessimistic, it is enough to check the property $\gamma(A) \leq \gamma^c(A)$ for focal sets. For if $A$ is not focal, then there is a focal set $E$ contained in $A$ such that $\gamma(A) = \gamma(E)$, and it is clear that if $\gamma(E) \leq \gamma^c(E)$ then $\gamma(A) \leq \gamma^c(E) \leq \gamma^c(A)$, since $A \subseteq E$. This remark helps checking the pessimism of $\gamma$.

**Proposition 2.** *The following properties hold:*

- *If $\gamma$ is pessimistic then $\gamma(A) = 1$ implies $\gamma(A^c) = 0$. For Boolean capacities, $\gamma$ is pessimistic if and only if* $\min(\gamma(A), \gamma(A^c)) = 0$.
- *If $\gamma$ is optimistic then $\gamma(A) = 0$ implies $\gamma(A^c) = 1$. For Boolean capacities, $\gamma$ is optimistic if and only if* $\max(\gamma(A), \gamma(A^c)) = 1$.

**Proof:** Let us suppose $\gamma$ pessimistic and $\gamma(A) = 1$. Then $\gamma^c(A) = 1$, hence $\gamma(A^c) = 0$. Conversely in the binary case, either $\gamma(A) = 0$ and then $\gamma(A) \leq \gamma^c(A)$, or $\gamma(A^c) = 0$ then $\gamma^c(A) = 1$ and $\gamma(A) \leq \gamma^c(A)$.

The proof for the optimistic case is similar. □

Note that we find a result proved in [2]. In that paper, a preference relation between acts represented by functions is called *strongly pessimistic* (resp. optimistic) if and only if it is represented by a Sugeno integral with respect to a necessity measure (resp. possibility measure). In that paper, contrary to ours, a preference relation is called pessimistic (resp. optimistic ) the relation is represented with a Sugeno integral with respect to capacity $\gamma$ such that $\min(\gamma(A), \gamma(A^c)) = 0$ (resp. $\max(\gamma(A), \gamma(A^c)) = 1$) [1]. Here we shall respectively call such capacities *strictly pessimistic* (resp. optimistic). Strictly pessimistic capacities are indeed special cases of pessimistic ones:

**Proposition 3.** *If a capacity is such that for all subsets $A$,* $\min(\gamma(A), \gamma(A^c)) = 0$ *then it is pessimistic.*

**Proof:** Suppose $\gamma$ is not pessimistic. Then $\exists A, \gamma(A) > \nu(\gamma(A^c))$. Hence $\gamma(A) > 0$, but then by assumption, $\gamma(A^c) = 0$, hence $\gamma(A) > 1$, which is impossible. □

[1] In the case of Boolean functions, Sugeno integral reduces to a capacity.

Considering a capacity $\gamma$, for each $\alpha > 0 \in L$ we can define a Boolean capacity $\gamma^\alpha$ called its $\alpha$-cut, as follows: for all $A \subseteq S$, $\gamma^\alpha(A)$ is equal to 1 if $\gamma(A) \geq \alpha$ and 0 otherwise. Then, $\gamma = \max_{\alpha \in L} \min(\alpha, \gamma^\alpha)$, $\mathcal{F}^\gamma = \cup_{\alpha \in L} \mathcal{F}^{(\gamma^\alpha)}$, and we can show:

**Proposition 4.** *A capacity is strictly pessimistic if and only if* $\forall \alpha > 0 \in L, \gamma^\alpha$ *is pessimistic.*

**Proof:** : If $\exists A, \alpha > 0 : \min(\gamma(A), \gamma(A^c)) = \alpha$, then $\gamma^\alpha(A) = \gamma^\alpha(A^c) = 1$ hence by Proposition 2, $\gamma^\alpha$ is not pessimistic. The converse is obvious. □

We can moreover describe the topological features of families of focal sets of pessimistic q-capacities:

**Proposition 5.** $\gamma$ *is a pessimistic q-capacity if and only if any two focal sets* $E$ *and* $F$ *that have empty intersection satisfy* $\gamma_\#(E) \leq \nu(\gamma_\#(F))$.

**Proof:** $\gamma \leq \gamma^c$ if and only if $\forall E \in \mathcal{F}^\gamma$, $\gamma_\#(E) \leq \nu(\gamma(E^c))$. But $\gamma(E^c) = \max_{F:F \cap E = \emptyset} \gamma_\#(F)$. So $\gamma_\#(E) \leq \min_{F:F \cap E = \emptyset} \nu(\gamma_\#(F))$. □

The condition $\gamma_\#(F) > \nu(\gamma_\#(E))$ that prevents disjoint focal sets means that weights $\gamma_\#(F), \gamma_\#(E)$ are both high enough. In particular, the focal sets $E$ of a pessimistic $\gamma$ such that $\gamma_\#(E) > \nu(\gamma_\#(E))$ (i.e., weights of focal sets in the upper part of the value scale) intersect pairwisely. All focal sets intersect the focal set with weight 1. And if the focal sets of a q-capacity all pairwisely intersect, then the q-capacity is pessimistic. This property is characteristic for Boolean capacities: $\gamma$ is a pessimistic Boolean capacity if and only if the intersection of two focal sets is not empty.

In the non-Boolean case, one may have disjoint focal sets with small enough weights inside or overlapping focal sets with high weights.

*Example 1.* Consider a q-capacity with 4 focal sets, $E, F, G_1, G_2$ such that $F \subseteq E$, $G_i \cap E \neq \emptyset$ for $i = 1, 2$, and $F, G_1, G_2$ are disjoint (Fig. 1). Assume $\gamma_\#(E) = 1, \gamma_\#(F) = \alpha > \nu(\alpha)$, and $\gamma_\#(G_i) = \beta_i < \nu(\alpha)$. Then it is easy to check that $\gamma$ is pessimistic, since $\gamma^c(E) = 1$, $\gamma^c(F) = \nu(\max(\beta_1, \beta_2)) > \alpha$, $\gamma^c(G_i) = \nu(\alpha) > \beta_i$.

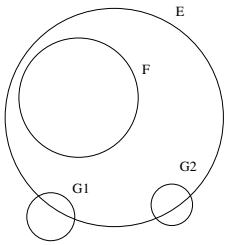



**Fig. 1.** Focal sets of a pessimistic q-capacity

For optimistic capacities one gets dual results:

**Proposition 6.** *Let $\gamma$ be an optimistic q-capacity. If $A$ is a focal set such that $\gamma(A) \neq 1$ then $A^c$ contains a focal set of $\gamma$.*

**Proof:** Let $A$ be a focal set of $\gamma$. $\gamma(A) = \alpha$ entails $\gamma(A^c) \geq \nu(\alpha) > 0$. □

So for each $A$ such that $\gamma_{\#}(A) \neq 1$ we can find another focal set $B$ such that $A \cap B = \emptyset$. It means that the focal sets of optimistic capacities tend to be disjoint while those of pessimistic capacities tend to overlap. The precise description of focal sets of optimistic capacities can benefit from the knowledge of the focal sets of pessimistic ones, as explained in [7].

## 5 Entailment between Capacities

Based on the above considerations, it makes sense to consider the following definition:

**Definition 2.** *If $\gamma_1$ and $\gamma_2$ are two pessimistic capacities then $\gamma_1$ is more informative than $\gamma_2$ if and only if $\gamma_1 \geq \gamma_2$ (eventwise).*

Indeed, if a pessimistic q-capacity $\gamma_1$ assigns a high value to event $A$, then it will assign a very small value to its complement. In particular $\gamma_1(A) = 1$ implies $\gamma_1(A^c) = 0$ (since $\nu(\gamma_1(A^c)) = 1$), so that this kind of capacity expresses the idea of certainty, while its conjugate expresses the notion of plausibility [5] and satisfies the opposite statement. Hence if $\gamma_1$ systematically assigns a certainty value higher than or equal to the one assigned by $\gamma_2$ to events, the former provides more information than the latter.

Note that by construction, $\gamma_1 \geq \gamma_2$ is equivalent to

$$[\gamma_1(A), \gamma_1^c(A)] \subseteq [\gamma_2(A), \gamma_2^c(A)], \forall A \subseteq S.$$

In fact $\gamma_1$ and its conjugate contain the same amount of information but differ by the attitude toward uncertainty. As pointed out in [5], the width of the interval $[\gamma_1(A), \gamma_1^c(A)]$ reflects the quantity of ignorance regarding event $A$, namely, for any pessimistic capacity

- $\gamma_1(A) = 1 (= \gamma_1^c(A))$ expresses the certainty of $A$.
- $\gamma_1^c(A) = 0 (= \gamma_1(A))$ expresses the impossibility of $A$.
- $\gamma_1^c(A) = 1$ and $\gamma_1(A) = 0$ expresses total ignorance regarding $A$.

Note that if the optimistic capacities are two possibility measures, this notion of relative information content reduces to the specificity ordering: $\Pi_1$ is more informative than $\Pi_2$ if and only if $\Pi_1 \leq \Pi_2$ if and only if $\pi_1$ is more specific than $\pi_2$, i.e. $\pi_1 \leq \pi_2$.

Given a capacity $\gamma$ one can derive its pessimistic and optimistic versions (respectively called assurance and opportunity functions by Yager [15]):

**Definition 3.** *The pessimistic version of a q-capacity $\gamma$ is $\gamma_*(A) = \min(\gamma(A), \gamma^c(A))$ and its optimistic version is $\gamma^*(A) = \max(\gamma(A), \gamma^c(A))$.*

Note that $\gamma_*$ and $\gamma^*$ are both q-capacities (monotonic increasing with inclusion), and, by construction, $\gamma^*$ is the $\nu$-conjugate of $\gamma_*$ (i.e. $\gamma^* = \gamma_*^c$) So each has the same information content as the other.

In such a context we can introduce a relation $\approx$ between q-capacities expressing the idea of containing the same amount of information :

**Definition 4.** *$\gamma$ and $\delta$ contain the same amount of information, denoted by $\gamma \approx \delta$ if and only if $\gamma^* = \delta^*$ and $\gamma_* = \delta_*$.*

This is an equivalence relation on the set of $L$-valued q-capacities on $S$. Note that we have $\gamma_* = \delta_*$ if and only if $\gamma^* = \delta^*$; so we just need to consider one equality. If $\gamma_* = \delta_*$ it means that for all subsets $A$ of $S$, the sets of values $\{\gamma(A), \gamma(A^c)\}$ and $\{\delta(A), \delta(A^c)\}$ are equal. So for each event $A$ we must decide the attitude in front of uncertainty : pessimistic if we assign the least value, optimistic if we assign the greatest one (while respecting monotonicity).

**Proposition 7.** *The equivalence class $\mathcal{C}_\approx(\gamma)$ of $\gamma$ is upper bounded by $\gamma^*$ and lower-bounded by $\gamma_*$.*

**Proof:** $\min(\gamma_*, \gamma_*^c) = \gamma_*$ so $\gamma_* \in \mathcal{C}_\approx(\gamma)$ and $\gamma_* \geq \min_{\delta \in \mathcal{C}_\approx(\gamma)} \delta$. Moreover for all $\delta \in \mathcal{C}_\approx(\gamma)$ we have $\delta \geq \min(\delta, \delta^c) = \delta_* = \gamma_*$ which entails $\min_{\delta \in \mathcal{C}_\approx(\gamma)} \delta \geq \gamma_*$. So $\min_{\delta \in \mathcal{C}_\approx(\gamma)} \delta = \gamma_*$. Similarly, we can prove that $\max_{\delta \in \mathcal{C}_\approx(\gamma)} \delta = \gamma^*$. □

If $\delta \in \mathcal{C}_\approx(\gamma)$ is a pessimistic q-capacity, then $\min(\delta, \delta^c) = \delta = \gamma_*$ so $\gamma_*$ is the unique pessimistic q-capacity in $\mathcal{C}_\approx(\gamma)$. Similarly one obtains that $\gamma^*$ is the unique optimistic q-capacity in $\mathcal{C}_\approx(\gamma)$. As a consequence, we must compare q-capacities in terms of their information content via a comparison of their equivalence classes, which justifies the following definition :

**Definition 5.** *A q-capacity $\gamma$ is said to be to be more informative than a q-capacity $\delta$ in the wide sense if and only if $\gamma_* \geq \delta_*$.*

Likewise we can compare two q-capacities in terms of their relative pessimism in front of uncertainty by means of another relation:

**Definition 6.** *A q-capacity $\gamma$ is said to be to be less pessimistic than a q-capacity $\delta$ in the wide sense if and only if $\{A : \gamma(A) = \gamma_*(A)\} \subseteq \{A : \delta(A) = \delta_*(A)\}$.*

These definitions completely disentangle the two aspects carried by q-capacities: namely the attitude in front of uncertainty and the information content, in agreement with possibilistic specificity.

## 6 Particular Cases

We consider in fact extreme cases of q-capacities.

**Complete Ignorance.** We consider the q-capacity defined by $\gamma(S) = 1$ and $\gamma(A) = 0$ otherwise. This is nothing but the vacuous necessity measure $N^?$ whose conjugate is $\Pi^?$. Clearly, $N^?$ is pessimistic and and it is less informative than any capacity, since $[N^?(A), \Pi^?(A)] = [0, 1]$ for $A \neq S, \emptyset$.

**Precise Information.** For each element $s_i$ in $S$ we define the following q-capacity $\sigma_i(A) = \begin{cases} 1 \text{ if } s_i \in A \\ 0 \text{ otherwise} \end{cases}$. It represents the precise information $x = s_i$. Note that $\sigma_i = \sigma_i^c$ is self-conjugate, so we have $\sigma_{i*} = \sigma_i^* = \sigma_i$. Some obvious remarks concerning this familly of q-capacities:

- If we consider two q-capacities $\sigma_i$ and $\sigma_j$ with $s_i \neq s_j$ then neither of the two is more informative than the other.
- $\sigma_i$ and $\sigma_j$ contain the same information if and only if $s_i = s_j$.
- $\mathcal{C}_{\approx}(\sigma_i) = \{\sigma_i\}$, which is obvious since it is self-conjugate.
- There is no pessimistic q-capacity $\gamma \neq \sigma_i$ such that $\gamma \geq \sigma_i$. Indeed $\sigma_i$ is self-conjugate and $[\sigma_{i*}(A), \sigma_i^*(A)]$ reduces to 0 or 1, i.e., cannot contain the non-trivial intervals $[\gamma_*(A), \gamma^*(A)]$. For each capacity $\gamma$, either the information content of $\gamma$ and $\sigma_i$ are incomparable or $\gamma$ is less informative than $\sigma_i$.

**Self-conjugate q-capacities.** In the numerical setting the most well-known self-conjugate capacities are probability measures that cannot be defined here, but for the precise q-capacities $\sigma_i$. Self-conjugate capacities $\sigma$ are such that $\forall A \subseteq S, \sigma(A) = \nu(\sigma(A^c)) = \sigma^c(A)$. They are at the same time pessimistic and optimistic. It is then obvious that they are maximally specific: there is no capacity that is more informative than any $\sigma$ since the intervals $[\sigma_*(A), \sigma^*(A)]$ are reduced to points, hence cannot contain $[\gamma_*(A), \gamma^*(A)]$ for any $\gamma \neq \sigma$. A remarkable subclass of self-conjugate capacities are symmetric ones, where $\gamma(A) = \alpha_{|A|}$ only depend on the cardinality of $A$, $\alpha_{|A|} = \nu(\alpha_{|S \setminus A|}) \geq \alpha_{|S \setminus A|}$ if $|A| \geq |S|/2$. They are completely defined by a strictly increasing sequence of $n = |S|$ coefficients $\alpha_i$ such that $\alpha_i = \nu(\alpha_{n-i+1})$. Of course, the precise q-capacities are a subclass of self-conjugate ones since if $\sigma_{\#}(\{s_i\}) = 1$ then $\mathcal{F}^{\sigma} = \{\{s_i\}\}$, i.e. $\sigma = \sigma_i$.

## 7 Conclusion

This paper has tried to define the information content of a capacity when the value scale is purely ordinal. We have exploited the fact that a qualitative capacity accounts for both uncertainty and attitude in front of it, as pointed out in [6,15]. We propose a mathematical definition of information content comparison irrespective of the attitude toward uncertainty, and a comparison of capacities in terms of pessimism, irrespective of how much uncertainty they express.

Some issues remain open, and especially the structure of the set of focal sets of optimistic capacities. Informally, the focal sets of a pessimistic capacity tend to overlap each other, while the focal sets of an optimistic capacity tend to be disjoint. And the larger the focal sets, the less informative the pessimistic capacity. A strict form of pessimism is when all focal sets intersect pairwisely; maximal pessimism is obtained for nested focal sets (necessity measures). Alternatively, maximal pessimism is obtained for disjoint focal sets (possibility measures); then the smaller the focal sets the less informative the optimistic capacity. It would

also be interesting to deepen the analogy and the differences between the qualitative and the quantitative settings for the modeling of uncertainty aversion. These results may be useful to clarify the potential use of qualitative capacities in various areas where aggregation functions are useful, especially in multicriteria analysis and information fusion. For instance, in a multiple source problem, q-capacities can be useful to merge possibility distributions [1]; focal sets of the q-capacity can then be understood as possibly antagonistic points of view on information.

## References


1. Assaghir, Z., Napoli, A., et al.: Numerical information fusion: lattice of answers with supporting arguments. In: Int. Conf. Tools for A.I., Boca Raton, pp. 621–628 (2011)
2. Chateauneuf, A., Grabisch, M., Rico, A.: Modeling attitudes toward uncertainty through the use of Sugeno integral. J. Math. Economics 44, 1084–1099 (2008)
3. Dubois, D., Prade, H.: A set-theoretic view of belief functions - Logical operations and approximation by fuzzy sets. Int. J. of General Systems 12(3), 193–226 (1986)
4. Dubois, D., Prade, H.: Qualitative possibility functions and integrals. In: Pap, E. (ed.) Handbook of Measure Theory, vol. 2, pp. 1469–1521. Elsevier, Amsterdam (2002)
5. Dubois, D., Prade, H.: Formal representations of uncertainty. In: Bouyssou, D., Dubois, D., Pirlot, M., Prade, H. (eds.) Decision-making Process- Concepts and Methods, ch. 3, pp. 85–156. ISTE London & Wiley (2009)
6. Dubois, D., Prade, H., Roubens, M., Sabbadin, R., Marichal, J.-L.: The use of the discrete Sugeno integral in decision-making: a survey. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems 9(5), 539–561 (2001)
7. Dubois, D., Prade, H., Rico, A.: Qualitative capacities as imprecise possibilities. In: van der Gaag, L.C. (ed.) ECSQARU 2013. LNCS, vol. 7958, pp. 169–180. Springer, Heidelberg (2013)
8. Dubois, D., Prade, H., Sabbadin, R.: Decision theoretic foundations of qualitative possibility theory. European Journal of Operational Research 128, 459–478 (2001)
9. Grabisch, M.: On the representation of $k$-decomposable measures. In: Proc. 7th IFSA world Congress, Prague, vol. 1, pp. 478–483 (1997)
10. Marichal, J.-L., Roubens, M.: Entropy of discrete fuzzy measures. Int. J. Uncert. Fuzziness, and Knowledge-Based Systems 8, 635–640 (2000)
11. Mesiar, R.: $k$-order pan-discrete fuzzy measures. In: Proc. 7th IFSA world Congress, Prague, vol. 1, pp. 488–490 (1997)
12. Prade, H., Rico, A.: Possibilistic evidence. In: Liu, W. (ed.) ECSQARU 2011. LNCS, vol. 6717, pp. 713–724. Springer, Heidelberg (2011)
13. Yager, R.R.: Entropy and specificity in a mathematical theory of evidence. International Journal of General Systems 9, 249–260 (1983)
14. Yager, R.R.: On the specificity of a possibility distribution. Fuzzy Sets and Systems 50, 279–292 (1992)
15. Yager, R.R.: Measures of assurance and opportunity in modeling uncertain information. Int. J. Intell. Syst. 27(8), 776–797 (2012)

# Maxitive Integral of Real-Valued Functions

Marco E.G.V. Cattaneo

Department of Statistics, LMU Munich
Ludwigstraße 33, 80539 München, Germany
cattaneo@stat.uni-muenchen.de
www.statistik.lmu.de/~cattaneo

**Abstract.** The paper pursues the definition of a maxitive integral on all real-valued functions (i.e., the integral of the pointwise maximum of two functions must be the maximum of their integrals). This definition is not determined by maxitivity alone: additional requirements on the integral are necessary. The paper studies the consequences of additional requirements of invariance with respect to affine transformations of the real line.

**Keywords:** maxitive measures, nonadditive integrals, location and scale invariance, Shilkret integral, convexity, subadditivity.

## 1 Introduction

In theories of reasoning and decision making under uncertainty, measures (or capacities) are used to describe uncertain belief or information, and can be extended to integrals in order to evaluate and compare uncertain (real-valued) payoffs. In particular, the additive capacities used in the Bayesian theory can be extended (almost) uniquely to an additive integral. By contrast, the extension to a maxitive integral of the maxitive capacities used in alternative theories is not unique, and additional requirements are needed in order to determine it. The present paper focuses on additional requirements of invariance with respect to the choice of the measurement scale of the payoffs.

The invariance with respect to the choice of the scale unit determines the Shilkret integral on nonnegative functions. This integral satisfies some properties that are important for the evaluation of uncertain payoffs, such as subadditivity or the law of iterated expectations, but it cannot be extended in a reasonable way to a maxitive integral on all real-valued functions. By contrast, the invariance with respect to the choice of the zero point of the measurement scale (almost) determines a maxitive integral on all real-valued functions, called convex integral. The name comes from the property of convexity, which is satisfied besides other important ones for the evaluation of uncertain payoffs, such as the law of iterated expectations.

The paper is organized as follows. The next section introduces the concepts of capacities and of integrals as their extensions. The Shilkret and convex integrals are then studied in Sects. 3 and 4, respectively. The final section gives directions for further research.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 226–235, 2014.

## 2 Integrals as Extensions of Capacities

Let $\Omega$ be a set and let $\mu$ be a capacity on $\Omega$. That is, $\mu : \mathcal{P}(\Omega) \to [0,1]$ is a monotonic set function such that $\mu(\varnothing) = 0$ and $\mu(\Omega) = 1$, where monotonic means that $\mu(A) \leq \mu(B)$ for all $A \subseteq B \subseteq \Omega$.

The capacity $\mu$ can be interpreted as a quantitative description of uncertain belief or information about $\omega \in \Omega$. The larger the value $\mu(A)$, the larger the plausibility of $\omega \in A$, or the larger the implausibility of $\omega \notin A$. This is in agreement with the monotonicity of $\mu$, while the requirements $\mu(\varnothing) = 0$ and $\mu(\Omega) = 1$ can be interpreted as meaning that $\omega \in \varnothing$ is impossible and that nothing speaks against $\omega \in \Omega$, respectively.

More precise interpretations of the values of $\mu$ can lead to additional requirements on the set function $\mu$. The best known additional requirement is (finite) additivity: $\mu(A \cup B) = \mu(A) + \mu(B)$ for all disjoint $A, B \subseteq \Omega$. Additive capacities (also called probability charges, or finitely additive probability measures) are the quantitative descriptions of uncertain belief about $\omega \in \Omega$ used in the Bayesian theory [1,2].

The continuity (from below and from above) of the set function $\mu$ is often required together with the additivity for technical reasons. The resulting requirement is countable additivity: $\mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mu(A_n)$ for all sequences $(A_n)_{n\in\mathbb{N}}$ of (pairwise) disjoint $A_n \subseteq \Omega$. Countably additive capacities (also called probability measures) are the quantitative descriptions of uncertain information about $\omega \in \Omega$ used in probability theory [3]. However, countable additivity is too strong a requirement when $\Omega$ is uncountable, and therefore $\mu$ cannot usually be defined on the whole power set $\mathcal{P}(\Omega)$, at least under the axiom of choice [4].

A common requirement on the set function $\mu$ alternative to additivity is (finite) maxitivity: $\mu(A \cup B) = \mu(A) \vee \mu(B)$ for all (disjoint) $A, B \subseteq \Omega$. Maxitive capacities (also called possibility measures, consonant plausibility functions, or idempotent measures) have been studied in various contexts [5,6,7]. As quantitative descriptions of uncertain belief or information about $\omega \in \Omega$, maxitive capacities play a central role in possibility theory [8], but they also appear for instance as consonant plausibility functions in the theory of belief functions [9], or as supremum preserving upper probabilities in the theory of imprecise probabilities [10]. Moreover, the description of uncertain belief by means of maxitive capacities also corresponds for example to the descriptions by means of degrees of potential surprise [11], or of degrees of support by eliminative induction [12]. Of particular importance in statistical applications is the fact that the likelihood (ratio) of composite hypotheses is a maxitive capacity [13,14].

The requirement of maxitivity of the set function $\mu$ can be generalized to the requirement of $\kappa$-maxitivity (where $\kappa$ is a cardinal): $\mu(\bigcup_{A\in\mathcal{A}} A) = \bigvee_{A\in\mathcal{A}} \mu(A)$ for all nonempty $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ of cardinality at most $\kappa$. Maxitivity corresponds to $\kappa$-maxitivity when $\kappa$ is finite and at least 2. Contrary to the case of additivity, the requirement of $\kappa$-maxitivity for infinite cardinals $\kappa$ does not pose any problem: a $\kappa$-maxitive set function can always be extended from a ring of subsets of $\Omega$ to the whole power set $\mathcal{P}(\Omega)$ [15, Theorem 1].

The capacity $\mu$ on $\Omega$ has a particularly simple description when it is $\kappa$-maxitive with $\kappa$ the cardinality of $\Omega$. In fact, $\mu$ is then completely described by its values on the singletons: $\mu(A) = \bigvee_{\omega\in A} \mu\{\omega\}$ for all nonempty $A \subseteq \Omega$. This implies in particular the $\kappa$-maxitivity of $\mu$ for all cardinals $\kappa$, also called complete maxitivity. For example, in statistics, the likelihood of composite hypotheses is a completely maxitive capacity: $\lambda(\mathcal{H}) = \bigvee_{\theta\in\mathcal{H}} \lambda\{\theta\}$ for all composite hypotheses $\mathcal{H} \subseteq \Theta$, where $\Theta$ is a set of simple hypotheses, and $\theta \mapsto \lambda\{\theta\}$ is the (relative) likelihood function on $\Theta$ [13,14].

Since each $A \subseteq \Omega$ can be identified with its indicator function $I_A$ on $\Omega$, the capacity $\mu$ can be identified with a functional on the set of all indicator functions $I_A$ with $A \subseteq \Omega$. The remaining of this paper studies extensions of this functional to larger classes of functions on $\Omega$, and in order to avoid trivial results, it is assumed that there is a $C \subseteq \Omega$ such that $0 < \mu(C) < 1$.

Let $\mathcal{F}$ be a set of extended real-valued functions $f : \Omega \to \overline{\mathbb{R}}$. A functional $F : \mathcal{F} \to \overline{\mathbb{R}}$ is said to extend the capacity $\mu$ to $\mathcal{F}$ if $F(I_A) = \mu(A)$ for all $A \subseteq \Omega$. In this definition, as in the rest of the paper, it is assumed as part of the condition that the expressions are well-defined. That is, $F$ can extend $\mu$ to $\mathcal{F}$ only if $I_A \in \mathcal{F}$ for all $A \subseteq \Omega$, because otherwise the expression $F(I_A) = \mu(A)$ would not be well-defined.

### 2.1 Extension of Additive Capacities

In the Bayesian theory, the uncertain belief about $\omega \in \Omega$ is described by an additive capacity $\mu$ on $\Omega$, while the evaluation of a (bounded) uncertain payoff $f(\omega)$ on the basis of this belief is given by its expectation $\int f \,\mathrm{d}\mu$, which is defined as follows.

Let $\mathcal{B}$ be the set of all bounded functions $f : \Omega \to \mathbb{R}$, and let $\mathcal{S} \subseteq \mathcal{B}$ be the subset of all simple functions (i.e., all functions $s : \Omega \to \mathbb{R}$ such that their images $s(\Omega)$ are finite). The (standard) integral of $f \in \mathcal{B}$ with respect to an additive capacity $\mu$ on $\Omega$ is denoted by $\int f \,\mathrm{d}\mu$ and is defined as

$$\int f \,\mathrm{d}\mu = \bigvee_{s\in\mathcal{S}\,:\,s\le f} \sum_{x\in s(\Omega)} x\,\mu\{s = x\} = \bigwedge_{s\in\mathcal{S}\,:\,s\ge f} \sum_{x\in s(\Omega)} x\,\mu\{s = x\},$$

where $\{s = x\}$ is the usual short form for the set $\{\omega \in \Omega : s(\omega) = x\}$. The integral is well-defined when $\mu$ is additive [16], and corresponds to the Lebesgue integral when $\mu$ is countably additive [17].

The next theorem shows that the integral with respect to an additive capacity $\mu$ on $\Omega$ is the unique monotonic, additive extension of $\mu$ to $\mathcal{B}$. A functional $F : \mathcal{F} \to \overline{\mathbb{R}}$ is said to be monotonic if $F(f) \le F(g)$ for all $f, g \in \mathcal{F}$ such that $f \le g$, while $F$ is said to be (finitely) additive if $F(f + g) = F(f) + F(g)$ for all $f, g \in \mathcal{F}$. Note that the additivity of a functional $F$ on $\mathcal{B}$ implies its monotonicity when some weak additional requirement is satisfied: for example when $F(f) \ge 0$ for all $f \in \mathcal{B}$ such that $f \ge 0$.

**Theorem 1.** *When $\mu$ is additive, its additive extension to $\mathcal{B}$ is not unique, but the functional $f \mapsto \int f \,\mathrm{d}\mu$ on $\mathcal{B}$ is the unique monotonic, additive extension of $\mu$ to $\mathcal{B}$.*

*Proof.* When $\mu$ is additive, the functional $F : f \mapsto \int f \,\mathrm{d}\mu$ extends $\mu$ and is monotonic and additive [16, Chap. 4]. If $F'$ is an additive extension of $\mu$ to $\mathcal{B}$, then its additivity implies $F'(\alpha\, f) = \alpha\, F'(f)$ for all $\alpha \in \mathbb{Q}$ and all $f \in \mathcal{B}$, and therefore also

$$F'(s) = \sum_{x \in s(\Omega)} x\, \mu\{s = x\} = F(s)$$

for all simple functions $s \in \mathcal{S}$ such that $s(\Omega) \subseteq \mathbb{Q}$. If $F'$ is also monotonic, then its monotonicity implies $F'(f) = F(f)$ for all $f \in \mathcal{B}$.

However, additive extensions of $\mu$ to $\mathcal{B}$ that are not monotonic also exist, at least under the axiom of choice. Let $\tau : \mathbb{R} \to \mathbb{R}$ be a discontinuous additive function such that $\tau(0) = 0$ and $\tau(1) = 1$ [18, Corollary 5.2.1]. Then the functional $F' : f \mapsto \int \tau \circ f \,\mathrm{d}\mu$ on $\mathcal{B}$ is an additive extension of $\mu$, but $F' \neq F$. □

### 2.2 Extension of Maxitive Capacities

In the Bayesian theory, the uncertain belief about $\omega \in \Omega$ is described by an additive capacity $\mu$ on $\Omega$, and the evaluations of uncertain payoffs $f \in \mathcal{B}$ are described by the unique monotonic, additive extension of $\mu$ to $\mathcal{B}$. Analogously, when the uncertain belief or information about $\omega \in \Omega$ is described by a maxitive capacity $\mu$ on $\Omega$, the evaluations of uncertain payoffs $f \in \mathcal{B}$ can be described by a maxitive extension of $\mu$ to $\mathcal{B}$. However, the next theorem shows that the maxitive extension to $\mathcal{B}$ of a maxitive capacity $\mu$ on $\Omega$ is not unique. A functional $F : \mathcal{F} \to \overline{\mathbb{R}}$ is said to be maxitive if $F(f \vee g) = F(f) \vee F(g)$ for all $f, g \in \mathcal{F}$. Note that the maxitivity of a functional implies its monotonicity.

**Theorem 2.** *When $\mu$ is maxitive, its maxitive extension to $\mathcal{B}$ is not unique.*

*Proof.* When $\mu$ is maxitive, both functionals

$$F : f \mapsto \bigvee_{x \in \mathbb{R}_{>0}} x\, \mu\{f > x\} \quad \text{and} \quad F' : f \mapsto \bigvee_{x \in \mathbb{R}_{>0}} (x \wedge \mu\{f > x\})$$

on $\mathcal{B}$ are maxitive extensions of $\mu$, because $\mu\{f \vee g > x\} = \mu\{f > x\} \vee \mu\{g > x\}$ for all $f, g \in \mathcal{B}$ and all $x \in \mathbb{R}$. However, $F \neq F'$, since for instance $F(2) = 2$, while $F'(2) = 1$. When $f \geq 0$, the values $F(f)$ and $F'(f)$ are known as Shilkret and Sugeno integrals of $f$ with respect to $\mu$, respectively [5,19]. □

In order to obtain a unique extension to $\mathcal{B}$ of a maxitive capacity $\mu$ on $\Omega$, additional requirements are necessary, besides maxitivity (and monotonicity). A particularly important requirement for evaluations of uncertain payoffs is their invariance with respect to changes in the measurement scale of the payoffs, such as changes in the location of the zero point or changes in the scale unit. A functional $F : \mathcal{F} \to \overline{\mathbb{R}}$ is said to be location invariant if $F(f + \alpha) = F(f) + \alpha$ for

all $f \in \mathcal{F}$ and all $\alpha \in \mathbb{R}$, while $F$ is said to be scale invariant if $F(\alpha\, f) = \alpha\, F(f)$ for all $f \in \mathcal{F}$ and all $\alpha \in \mathbb{R}_{>0}$.

The (standard) integral with respect to additive capacities is location and scale invariant [16]. The best known location and scale invariant integral with respect to nonadditive capacities is the one of Choquet [20,21]. The Choquet integral of $f \in \mathcal{B}$ with respect to a capacity $\mu$ on $\Omega$ is denoted by $\int^{\mathrm{C}} f \,\mathrm{d}\mu$ and is defined as

$$\int^{\mathrm{C}} f \,\mathrm{d}\mu = \int_{-\infty}^{0} (\mu\{f > x\} - 1)\,\mathrm{d}x + \int_{0}^{+\infty} \mu\{f > x\}\,\mathrm{d}x,$$

where the right-hand side is the well-defined sum of two improper Riemann integrals. The Choquet integral with respect to a capacity $\mu$ on $\Omega$ is a monotonic extension of $\mu$ to $\mathcal{B}$, which is additive when $\mu$ is additive [22]. Therefore, $\int^{\mathrm{C}} f \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$ for all $f \in \mathcal{B}$ when $\mu$ is additive.

The next theorem shows that no maxitive extension to $\mathcal{B}$ of a maxitive capacity $\mu$ on $\Omega$ is location and scale invariant. Maxitive extensions satisfying one of these two additional requirements are studied in the next two sections.

**Theorem 3.** *When $\mu$ is maxitive, there is no location and scale invariant, maxitive extension of $\mu$ to $\mathcal{B}$.*

*Proof.* Let $F$ be a scale invariant, maxitive extension to $\mathcal{B}$ of a maxitive capacity $\mu$ on $\Omega$. As assumed above, there is a $C \subseteq \Omega$ such that $0 < \mu(C) < 1$. Hence, $\mu(\Omega \setminus C) = 1$ and

$$F(I_C + 1) = F\left((2\, I_C) \vee I_{\Omega \setminus C}\right) = (2\,\mu(C)) \vee 1 < \mu(C) + 1 = F(I_C) + 1,$$

and therefore $F$ is not location invariant. □

## 3 Shilkret Integral

Let $\mathcal{E}$ be the set of all extended real-valued functions $f : \Omega \to \overline{\mathbb{R}}$, let $\mathcal{E}^+ \subseteq \mathcal{E}$ be the subset of all nonnegative functions, and let $\mathcal{B}^+ = \mathcal{B} \cap \mathcal{E}^+$ the subset of all bounded, nonnegative functions. The Shilkret integral of $f \in \mathcal{E}^+$ with respect to a capacity $\mu$ on $\Omega$ is denoted by $\int^{\mathrm{S}} f \,\mathrm{d}\mu$ and is defined as

$$\int^{\mathrm{S}} f \,\mathrm{d}\mu = \bigvee_{x \in \mathbb{R}_{>0}} x\,\mu\{f > x\}.$$

The Shilkret integral has a particularly simple expression when $\mu$ is completely maxitive: $\int^{\mathrm{S}} f \,\mathrm{d}\mu = \bigvee_{\omega \in \Omega} f(\omega)\,\mu\{\omega\}$ for all $f \in \mathcal{E}^+$ [5].

The next theorem shows that the Shilkret integral with respect to a maxitive capacity $\mu$ on $\Omega$ is the unique scale invariant, maxitive extension of $\mu$ to $\mathcal{B}^+$. The Shilkret integral maintains $\kappa$-maxitivity also for infinite cardinals $\kappa$. A functional $F : \mathcal{F} \to \overline{\mathbb{R}}$ is said to be $\kappa$-maxitive if $F(\bigvee_{f \in \mathcal{G}} f) = \bigvee_{f \in \mathcal{G}} F(f)$ for all nonempty $\mathcal{G} \subseteq \mathcal{F}$ of cardinality at most $\kappa$.

**Theorem 4.** *When $\mu$ is maxitive, the functional $f \mapsto \int^{\mathrm{S}} f \,\mathrm{d}\mu$ on $\mathcal{B}^+$ is the unique scale invariant, maxitive extension of $\mu$ to $\mathcal{B}^+$. Moreover, when $\kappa$ is an infinite cardinal and $\mu$ is $\kappa$-maxitive, the functional $f \mapsto \int^{\mathrm{S}} f \,\mathrm{d}\mu$ on $\mathcal{E}^+$ is the unique scale invariant, $\kappa$-maxitive extension of $\mu$ to $\mathcal{E}^+$.*

*Proof.* When $\kappa$ is a cardinal and $\mu$ is $\kappa$-maxitive, the functional $f \mapsto \int^{\mathrm{S}} f \,\mathrm{d}\mu$ on $\mathcal{E}^+$ is a scale invariant, $\kappa$-maxitive extension of $\mu$ to $\mathcal{E}^+$ [15, Lemma 1]. Such an extension is unique on $\mathcal{B}^+$ when $\kappa \geq 2$ [15, Theorem 2 (iii)], and it is unique also on $\mathcal{E}^+$ when $\kappa$ is infinite [15, Theorem 3 (iii)]. □

An important property for evaluations of uncertain payoffs is convexity, meaning that diversification does not increase the risk [23,24]. A functional $F : \mathcal{F} \to \overline{\mathbb{R}}$ is said to be convex if $F(\lambda f + (1-\lambda) g) \leq \lambda F(f) + (1-\lambda) F(g)$ for all $\lambda \in (0,1)$ and all $f, g \in \mathcal{F}$, whereas $F$ is said to be subadditive if $F(f+g) \leq F(f) + F(g)$ for all $f, g \in \mathcal{F}$. Note that convexity and subadditivity are equivalent for a scale invariant functional.

The characterization of the capacities with respect to which the Choquet integral is convex (i.e., subadditive) is a well-known result [20,22,21]. The next theorem gives also a characterization of the capacities with respect to which the Shilkret integral is convex (i.e., subadditive). The capacity $\mu$ is said to be submodular if $\mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B)$ for all $A, B \subseteq \Omega$. Note that additive or maxitive capacities are submodular.

**Theorem 5.** *The functional $f \mapsto \int^{\mathrm{C}} f \,\mathrm{d}\mu$ on $\mathcal{B}$ is convex if and only if $\mu$ is submodular, while the functional $f \mapsto \int^{\mathrm{S}} f \,\mathrm{d}\mu$ on $\mathcal{E}^+$ is convex if and only if $\mu$ is maxitive.*

*Proof.* Both functionals $f \mapsto \int^{\mathrm{C}} f \,\mathrm{d}\mu$ on $\mathcal{B}$ and $f \mapsto \int^{\mathrm{S}} f \,\mathrm{d}\mu$ on $\mathcal{E}^+$ are scale invariant. The first one is subadditive if and only if $\mu$ is submodular [21, Chap. 6], while the second one is subadditive if and only if $\mu$ is maxitive [15, Theorem 4 (iii)]. □

The Shilkret integral with respect to maxitive capacities satisfies also other important properties for evaluations of uncertain payoffs, such as the law of iterated expectations (or evaluations). By contrast, the Choquet integral satisfies this law only with respect to additive capacities (i.e., only when it corresponds to the standard integral) [15,25]. However, the Shilkret integral is defined only for nonnegative functions, and its extension to functions taking negative values is problematic.

When $\mu$ is maxitive, the Shilkret integral is the unique scale invariant, maxitive extension of $\mu$ to $\mathcal{B}^+$, but the next theorem shows that its further (scale invariant, maxitive) extension to $\mathcal{B}$ is not unique. In fact, the values assigned to negative functions by such extensions are independent from $\mu$. To impose a dependence from $\mu$, some kind of symmetry of the extension could be required. For example, since $\mu$ determines the values of its extensions on all indicator functions $I_A$, the determination of the values on all negative indicator functions $-I_A$ could be required. An extension $F : \mathcal{F} \to \overline{\mathbb{R}}$ of a capacity $\mu$ on $\Omega$ is said

to be symmetric if $F(-I_A) = -\mu(A)$ for all $A \subseteq \Omega$, while $F$ is said to be dual symmetric if $F(-I_A) = -\overline{\mu}(A)$ for all $A \subseteq \Omega$, where the dual capacity $\overline{\mu}$ on $\Omega$ is defined by $\overline{\mu}(A) = 1 - \mu(\Omega \setminus A)$. Note that all location invariant extensions of a capacity are dual symmetric, and the (standard) integral with respect to an additive capacity is also symmetric.

However, the next theorem also shows that no scale invariant, maxitive extension to $\mathcal{B}$ of a maxitive capacity $\mu$ on $\Omega$ is symmetric or dual symmetric, and neither is it convex and calibrated. A functional $F : \mathcal{F} \to \overline{\mathbb{R}}$ is said to be calibrated if $F(\alpha) = \alpha$ for all constant functions $\alpha \in \mathcal{F}$. Note that all scale invariant extensions of $\mu$ to $\mathcal{B}^+$, all scale invariant, (dual) symmetric extensions of $\mu$ to $\mathcal{B}$, and all location invariant extensions of $\mu$ to $\mathcal{B}$ are calibrated.

**Theorem 6.** *When $\mu$ is maxitive and $\gamma$ is a real-valued function on $\Omega$ such that $\bigwedge_{\omega\in\Omega} \gamma(\omega) = 1$, the functional*

$$f \mapsto \begin{cases} \bigvee_{\omega\in\Omega} f(\omega)\,\gamma(\omega) & \text{if } f < 0, \\ \int^{\mathrm{S}} (f \vee 0)\,\mathrm{d}\mu & \text{otherwise} \end{cases}$$

*on $\mathcal{E}$ is a scale invariant, calibrated, maxitive extension of $\mu$ to $\mathcal{E}$, but no scale invariant, calibrated, maxitive extension of $\mu$ to $\mathcal{B}$ is symmetric, dual symmetric, or convex.*

*Proof.* Since the functional $f \mapsto \int^{\mathrm{S}} f\,\mathrm{d}\mu$ on $\mathcal{E}^+$ is a scale invariant, calibrated, maxitive extension of $\mu$ to $\mathcal{E}^+$ when $\mu$ is maxitive, its further extension to $\mathcal{E}$ defined in the theorem is also scale invariant, calibrated, and maxitive.

Let $F$ be a scale invariant, calibrated, maxitive extension to $\mathcal{B}$ of a maxitive capacity $\mu$ on $\Omega$. As assumed above, there is a $C \subseteq \Omega$ such that $0 < \mu(C) < 1$. Hence, $\mu(\Omega \setminus C) = 1$ and

$$F(-I_C) \vee F(-I_{\Omega\setminus C}) = 0 > -\mu(C) = (-\mu(C)) \vee (-\mu(\Omega \setminus C))\,,$$

and therefore $F$ is not symmetric. Neither can $F$ be dual symmetric, because $0 < \overline{\mu}(\Omega \setminus C) < 1$, while

$$F(-I_{\Omega\setminus C}) = F\left((-2\,I_{\Omega\setminus C}) \vee (-1)\right) = \left(2\,F(-I_{\Omega\setminus C})\right) \vee (-1)$$

implies $F(-I_{\Omega\setminus C}) \in \{-1, 0\}$. Finally, if $f = I_C \vee \mu(C)$, then $F(f) = \mu(C)$ and since

$$F\left(f + (-\mu(C))\right) = F\left((1 - \mu(C))\ I_C\right) > 0 = F(f) + F\left(-\mu(C)\right),$$

$F$ is not subadditive (i.e., convex). □

## 4 Convex Integral

The convex integral of $f \in \mathcal{E}$ with respect to a capacity $\mu$ on $\Omega$ is denoted by $\int^{\mathrm{X}} f\,\mathrm{d}\mu$ and is defined as

$$\int^{\mathrm{X}} f\,\mathrm{d}\mu = \bigvee_{x\in\mathbb{R}} \left(x + \tau \circ \mu\{f > x\}\right),$$

where $\tau$ is the function on $[0,1]$ defined by $\tau(0) = -\infty$ and $\tau(x) = x - 1$ otherwise. The convex integral has a particularly simple expression when $\mu$ is completely maxitive: $\int^{\mathrm{X}} f \,\mathrm{d}\mu = \bigvee_{\omega\in\Omega\,:\,\mu\{\omega\}>0} (f(\omega) + \mu\{\omega\} - 1)$ for all $f \in \mathcal{E}$.

The next theorem shows that the convex integral with respect to a maxitive capacity $\mu$ on $\Omega$ is the unique location invariant, maxitive extension of $\mu$ to $\mathcal{B}$, when $\varnothing$ is the only null set (i.e., $\mu(A) > 0$ for all nonempty $A \subseteq \Omega$). When there are nonempty null sets, the location invariant, maxitive extension to $\mathcal{B}$ of a maxitive capacity $\mu$ on $\Omega$ is not unique, but the convex integral is the only null preserving one. An extension $F : \mathcal{F} \to \overline{\mathbb{R}}$ of a capacity $\mu$ on $\Omega$ is said to be null preserving if $F(f) = 0$ for all $f \in F$ such that $\mu\{f \neq 0\} = 0$. Note that all extensions of a capacity are null preserving when $\varnothing$ is the only null set.

**Theorem 7.** *When $\mu$ is maxitive, the functional $f \mapsto \int^{\mathrm{X}} f \,\mathrm{d}\mu$ on $\mathcal{B}$ is the unique location invariant, maxitive extension of $\mu$ to $\mathcal{B}$ if and only if $\varnothing$ is the only null set, and in general it is the unique location invariant, null preserving, maxitive extension of $\mu$ to $\mathcal{B}$. Moreover, when $\kappa$ is an infinite cardinal and $\mu$ is $\kappa$-maxitive, the functional $f \mapsto \int^{\mathrm{X}} f \,\mathrm{d}\mu$ on $\mathcal{E}$ is the unique location invariant, null preserving, $\kappa$-maxitive extension of $\mu$ to $\mathcal{E}$.*

*Proof.* When $\kappa \geq 2$ is a cardinal and $\mu$ is $\kappa$-maxitive, the functional $f \mapsto \int^{\mathrm{X}} f \,\mathrm{d}\mu$ on $\mathcal{E}$ is a location invariant, null preserving, $\kappa$-maxitive extension of $\mu$ to $\mathcal{E}$ [15, Corollary 5]. Such an extension is unique on $\mathcal{B}$ [15, Corollary 6], and it is unique also on $\mathcal{E}$ when $\kappa$ is infinite [15, Corollary 7].

Let $\nu$ be the set function on $\mathcal{P}(\Omega)$ defined by $\nu(\varnothing) = -\infty$ and $\nu(A) = \mu(A) - 1$ otherwise. When $\mu$ is maxitive, the functional $f \mapsto \bigvee_{x\in\mathbb{R}} (x + \nu\{f > x\})$ on $\mathcal{B}$ is a location invariant, maxitive extension of $\mu$ to $\mathcal{B}$ [15, Corollary 6], and it differs from the functional $f \mapsto \int^{\mathrm{X}} f \,\mathrm{d}\mu$ on $\mathcal{B}$ when there are nonempty null sets. □

Convexity and subadditivity are not equivalent for functionals that are not scale invariant. The next theorem shows that the convex integral with respect to maxitive capacities is not subadditive. However, it is convex, and this is the reason for its name.

**Theorem 8.** *The functional $f \mapsto \int^{\mathrm{X}} f \,\mathrm{d}\mu$ on $\mathcal{B}$ is convex if and only if $\mu$ is maxitive. But when $\mu$ is maxitive, no location invariant, maxitive extension of $\mu$ to $\mathcal{B}$ is subadditive.*

*Proof.* When $\mu$ is maxitive, the functional $f \mapsto \int^{\mathrm{X}} f \,\mathrm{d}\mu$ on $\mathcal{B}$ is convex [15, Theorem 7]. But when $\mu$ is not maxitive, there are $A, B \subseteq \Omega$ and $\alpha \in \mathbb{R}_{>0}$ such that $\mu(A \cup B) - \alpha > \mu(A) \vee \mu(B)$. Hence, if $g = I_{A\cup B} + \alpha\, I_B$ and $h = I_{A\cup B} - \alpha\, I_B$, then $\int^{\mathrm{X}} g \,\mathrm{d}\mu = \mu(A \cup B)$ and $\int^{\mathrm{X}} h \,\mathrm{d}\mu = \mu(A \cup B) - \alpha$, and since

$$\int^{\mathrm{X}} \left(\tfrac{1}{2}\, g + \tfrac{1}{2}\, h\right) \mathrm{d}\mu = \mu(A \cup B) > \tfrac{1}{2} \int^{\mathrm{X}} g \,\mathrm{d}\mu + \tfrac{1}{2} \int^{\mathrm{X}} h \,\mathrm{d}\mu,$$

the functional $f \mapsto \int^{\mathrm{X}} f \,\mathrm{d}\mu$ on $\mathcal{B}$ is not convex.

Let $F$ be a location invariant, maxitive extension to $\mathcal{B}$ of a maxitive capacity $\mu$ on $\Omega$. As assumed above, there is a $C \subseteq \Omega$ such that $0 < \mu(C) < 1$, and thus there is an $n \in \mathbb{N}$ such that $n\ (1 - \mu(C)) \geq 1$. Hence, if $f = (1 - \mu(C))\ I_C$, then

$$F(n\, f) \geq \mu(C) > 0 = n\ (F\,(I_C \vee \mu(C)) - \mu(C)) = n\, F(f),$$

and therefore $F$ is not subadditive. □

Besides convexity and location invariance, the convex integral with respect to maxitive capacities satisfies also other important properties for evaluations of uncertain payoffs, such as the law of iterated expectations (or evaluations) [15]. The convex integral can be generalized by replacing the set function $\tau \circ \mu$ in its definition with an arbitrary monotonic set function $\nu$ on $\mathcal{P}(\Omega)$ such that $\nu(\varnothing) = -\infty$ and $\nu(\Omega) = 0$, also called a penalty on $\Omega$ [15].

In particular, the convex integral with respect to completely maxitive capacities (or penalties) is strictly related to the idempotent integral of tropical mathematics [7] and to convex measures of risk [24]. It corresponds to the functional $f \mapsto \bigvee_{\omega \in \Psi} (f(\omega) - \psi(\omega))$ on $\mathcal{E}$, where $\Psi \subseteq \Omega$ is not empty and $\psi$ is a real-valued function on $\Psi$ such that $\bigwedge_{\omega \in \Psi} \psi(\omega) = 0$.

## 5 Conclusion

The present paper studied maxitive integrals with respect to maxitive capacities, and in particular the Shilkret and convex integrals. These have particularly simple expressions when the capacities are completely maxitive. In this case, the Shilkret and convex integrals can be characterized as evaluations of uncertain payoffs by few basic decision-theoretic properties. These will be discussed in future work, with particular emphasis on the case of likelihood-based decision making [14,26].

**Acknowledgments.** The author wishes to thank an anonymous referee for valuable comments and suggestions.

## References

1. de Finetti, B.: Theory of Probability, 2 vols. Wiley (1974-1975)
2. Savage, L.J.: The Foundations of Statistics, 2nd edn. Dover (1972)
3. Kolmogorov, A.N.: Foundations of the Theory of Probability, 2nd edn. Chelsea Publishing (1956)
4. Solovay, R.M.: A model of set-theory in which every set of reals is Lebesgue measurable. Ann. Math. (2) 92, 1–56 (1970)
5. Shilkret, N.: Maxitive measure and integration. Indag. Math. 33, 109–116 (1971)
6. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. 1, 3–28 (1978)
7. Kolokoltsov, V.N., Maslov, V.P.: Idempotent Analysis and Its Applications. Springer (1997)

8. Dubois, D., Prade, H.: Possibility Theory. Plenum Press (1988)
9. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
10. de Cooman, G., Aeyels, D.: Supremum preserving upper probabilities. Inf. Sci. 118, 173–212 (1999)
11. Shackle, G.L.S.: Expectation in Economics. Cambridge University Press (1949)
12. Cohen, L.J.: A logic for evidential support. Br. J. Philos. Sci. 17, 21–43, 105–126 (1966)
13. Neyman, J., Pearson, E.S.: On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika 20A, 175–240, 263–294 (1928)
14. Cattaneo, M.: Likelihood decision functions. Electron. J. Stat. 7, 2924–2946 (2013)
15. Cattaneo, M.: On maxitive integration. Technical Report 147, Department of Statistics, LMU Munich (2013)
16. Bhaskara Rao, K.P.S., Bhaskara Rao, M.: Theory of Charges. Academic Press (1983)
17. Ash, R.B.: Real Analysis and Probability. Academic Press (1972)
18. Kuczma, M.: An Introduction to the Theory of Functional Equations and Inequalities, 2nd edn. Birkhäuser (2009)
19. Sugeno, M.: Theory of Fuzzy Integrals and Its Applications. PhD thesis, Tokyo Institute of Technology (1974)
20. Choquet, G.: Theory of capacities. Ann. Inst. Fourier 5, 131–295 (1954)
21. Denneberg, D.: Non-Additive Measure and Integral. Kluwer (1994)
22. Schmeidler, D.: Integral representation without additivity. Proc. Am. Math. Soc. 97, 255–261 (1986)
23. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Finance 9, 203–228 (1999)
24. Föllmer, H., Schied, A.: Stochastic Finance, 3rd edn. De Gruyter (2011)
25. Yoo, K.R.: The iterative law of expectation and non-additive probability measure. Econ. Lett. 37, 145–149 (1991)
26. Giang, P.H., Shenoy, P.P.: Decision making on the sole basis of statistical likelihood. Artif. Intell. 165, 137–163 (2005)

# Fuzzy Weber Sets and Lovász Extensions of Cooperative Games

Nobusumi Sagara*

Faculty of Economics, Hosei University
4342, Aihara, Machida, Tokyo, 194–0298 Japan
nsagara@hosei.ac.jp

**Abstract.** This paper investigates fuzzy extensions of cooperative games and the coincidence of the solutions for fuzzy and crisp games. We first show that an exact game has an exact fuzzy extension such that its fuzzy core coincides with the core. For games with empty cores, we exploit Lovász extensions to establish the coincidence of Weber sets for fuzzy and crisp games.

**Keywords:** Core, Weber set, Exact game, Lovász extension, Regularity, Convex games.

## 1 Introduction

Classical cooperative game theory deals with the situation where players have only two alternative possibilities. That is, whether they join a coalition or not, without any option for the degree of commitment. Further, once the players join a coalition, all of them are required to commit themselves fully to the coalition. On the contrary, cooperative fuzzy games proposed by [1, 2] allow for the partial participation of players in coalitions where the attainable outcomes of a game depends on the degree of commitment of the players, thereby modelling ambiguous decision making as observed in most group behavior.

Fuzzy games are defined on fuzzy coalitions. The restriction of fuzzy games to usual coalitions (characteristic functions) yields a crisp (or nonfuzzy) game. However, there are numerous fuzzy games that are fuzzy extensions of a given game and there are innumerable fuzzy games that yield the same crisp game. Similar to classical cooperative game theory, the fuzzy core is a fundamental solution concept in fuzzy game theory. Since the fuzzy core of a fuzzy game is included in the core of its original (or crisp) game, part of the information on the core of a crisp game is inevitably lost in the fuzzy core. This means that there exists a payoff vector in the core of a crisp game that is blocked by the forming of a fuzzy coalition. Can such a payoff vector be a plausible solution of a game? Importantly, the core of a crisp game is a reasonable solution concept only when the crisp game does not permit fuzzification.

* This research is supported by a Grant-in-Aid for Scientific Research (No. 23530230) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 236–245, 2014.

The first question addressed in this paper is as follows. Under what conditions in crisp games and their fuzzy extensions is the coincidence of the core of a fuzzy game and that of a crisp game guaranteed? The response given in the paper is that consistent classes of games and their fuzzy extensions that guarantee the core coincidence is the class of exact games and their exact extensions investigated in [12, 15, 16]. For the core coincidence, the useful observation made by [1] that the superdifferential of a positively homogeneous fuzzy game is the fuzzy core plays a crucial role because the above extensions are positively homogeneous, concave functions and the standard convex analysis are effectively used.

However, if the core of a crisp game is empty, which can happen in many applications, then the core coincidence is utterly in vain. In this case, what kind of solutions and fuzzy extensions should be pursued for the coincidence of the solutions for fuzzy and crisp games? This is the second question addressed in this paper. To answer this natural question, we propose to exploit Weber sets and Lovász extensions (or Choquet integrals) for the class of "all" crisp games. The advantage of Weber sets is exemplified by the fact that the Weber set of any crisp game is nonempty and contains the core (see [19]).

The Weber set is defined as the convex hull of the marginal contributions of each player. We define the fuzzy Weber set of a fuzzy game via the limiting marginal contributions of each player in terms of the directional derivative of a fuzzy game and demonstrate that the fuzzy Weber set of the Lovász extension of any crisp game coincides with the Weber set. Toward this end, we prove that the Clarke superdifferential of the Lovász extension is the fuzzy Weber set. The fact that Lovász extensions are not necessarily concave or differentiable furnishes one with another reason to utilize the Clarke superdifferential, a generalized notion of the superdifferentials in convex analysis (see [5]), to investigate Weber sets. Based on the powerful technique of nonsmooth analysis brought into cooperative game theory, we provide an additional characterization of convex games in terms of the regularity of the Lovász extensions, as a weaker form of smoothness.

## 2 Preliminaries

### 2.1 Clarke Superdifferentials

The *directional derivative* of an extended real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ at $x \in \mathbb{R}^n$ in the direction $h \in \mathbb{R}^n$ is defined by

$$f'(x;h) = \lim_{\lambda \downarrow 0} \frac{f(x+\lambda h) - f(x)}{\lambda}$$

when this limit exists in $\mathbb{R}$. The *superdifferential* $\partial f(x)$ of $f$ at $x$ is given by

$$\partial f(x) = \{p \in \mathbb{R}^n \mid f(y) - f(x) \leq \langle p, y - x\rangle \ \forall y \in \mathbb{R}^n\},$$

where we denote by $\langle x, y\rangle$ the inner product of the vectors $x, y \in \mathbb{R}^n$. An element in $\partial f(x)$ is called a *supergradient* of $f$ at $x$. It follows from the standard argument

of convex analysis in finite dimensions that if $f$ is concave and $f(x)$ is finite, then $\partial f(x)$ is a nonempty, compact, convex subset of $\mathbb{R}^n$, the directional derivative $f'(x;h)$ exists in any direction $h \in \mathbb{R}^n$,

$$\partial f(x) = \{p \in \mathbb{R}^n \mid f'(x;h) \leq \langle p, h\rangle \ \forall h \in \mathbb{R}^n\}$$

and

$$f'(x;h) = \min_{p \in \partial f(x)} \langle p, h\rangle \quad \text{for every } h \in \mathbb{R}^n. \tag{2.1}$$

(See [14, Theorems 23.1, 23.2 and 23.4].)

A real-valued function $f : \mathbb{R}^n \to \mathbb{R}$ is *Lipschitz of rank* $K \geq 0$ near a given point $\bar{x} \in \mathbb{R}^n$ if there exists some $\varepsilon > 0$ such that

$$|f(x) - f(y)| \leq K\|x - y\| \quad \text{for every } x, y \in \bar{x} + \varepsilon B.$$

Here, $B$ is the open unit ball in $\mathbb{R}^n$ and $\|\cdot\|$ is the Euclid norm of $\mathbb{R}^n$. A function $f$ is said to be *locally Lipschitz* if $f$ is Lipschitz near $x$ for every $x \in \mathbb{R}^n$. Let $f$ be Lipschitz near $x$. The *Clarke directional derivative* of $f$ at $x$ in the direction $h \in \mathbb{R}^n$, denoted by $f^\circ(x;h)$, is defined as follows

$$f^\circ(x;h) = \liminf_{\substack{y \to x \\ \lambda \downarrow 0}} \frac{f(y + \lambda h) - f(y)}{\lambda}.$$

The *Clarke superdifferential* of $f$ at $x$, denoted by $\partial_{\mathrm{C}} f(x)$, is defined by

$$\partial_{\mathrm{C}} f(x) = \{p \in \mathbb{R}^n \mid f^\circ(x;h) \leq \langle p, h\rangle \ \forall h \in \mathbb{R}^n\}.$$

If $f$ is Lipschitz near $x$, then $\partial_{\mathrm{C}} f(x)$ is nonempty, convex and compact. Furthermore,

$$f^\circ(x;h) = \min_{p \in \partial_{\mathrm{C}} f(x)} \langle p, h\rangle \quad \text{for every } h \in \mathbb{R}^n. \tag{2.2}$$

(See [5, Proposition 2.1.2]). The function $f$ is *regular* at $x$ provided that $f$ admits the directional derivative $f'(x;h)$ at $x$ with $f'(x;h) = f^\circ(x;h)$ for every $h \in \mathbb{R}^n$. If $f$ is concave, then it is regular at every point $x$ and $\partial_{\mathrm{C}} f(x) = \partial f(x)$ (see [5, Propositions 2.2.6 and 2.2.7]).

### 2.2 Cores, Exact Games and Weber Sets

Denote by $N = \{1, \ldots, n\}$ the finite set of *players* with its generic element denoted by $i$. A nonempty subset of $N$ is called a *coalition*. A set function $v : 2^N \to \mathbb{R}$ with $v(\emptyset) = 0$ is a *game*, which is alternatively denoted by $(N, v)$. A *payoff vector* is an element in $\mathbb{R}^n$. For any payoff vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and coalition $S \in 2^N$, define $x(S) = \sum_{i \in S} x_i$ and $x(\emptyset) = 0$. A payoff vector $x \in \mathbb{R}^n$ is *feasible* for a game $v$ if $x(N) \leq v(N)$. The set $\mathcal{C}(v)$ of feasible payoff vectors which cannot be improved upon by any coalition is the *core* of a game $v$ is defined by

$$\mathcal{C}(v) = \{x \in \mathbb{R}^n \mid x(N) = v(N),\ x(S) \geq v(S)\ \forall S \in 2^N\}.$$

The characteristic function of $S \in 2^N \setminus \{\emptyset\}$ is denoted by $\mathbb{1}_S$, which is identified with a vector in $\mathbb{R}^n$ such that its $i$-th component is 1 if $i \in S$ and 0 otherwise. A game is *balanced* if its core is nonempty. A balanced game $v$ is *exact* if

$$v(S) = \min_{x \in \mathcal{C}(v)} x(S) \quad \text{for every } S \in 2^N.$$

A game $v$ is *convex* (or *supermodular*) if

$$v(S) + v(T) \leq v(S \cup T) + v(S \cap T) \quad \text{for every } S, T \in 2^N.$$

Convex games are exact. Convex games are exact (see [4, 16]).

Denote by $\Pi_N$ the set of all permutations on $N$. For each $i \in N$, define the subset of $N$ by

$$P_i^\pi = \{j \in N \mid \pi(j) < \pi(i)\}.$$

Then $P_i^\pi$ is the set of players precedent to $i$ with respect to the order $\pi \in \Pi_N$. Define the payoff vector $a^\pi(v) \in \mathbb{R}^n$ whose component is the *marginal contribution* of each player by

$$a_i^\pi(v) = v(P_i^\pi \cup \{i\}) - v(P_i^\pi), \quad i \in N.$$

The *Weber set* of a game $v$ is defined by

$$\mathcal{W}(v) = \mathrm{co}\{a^\pi(v) \in \mathbb{R}^n \mid \pi \in \Pi_N\}.$$

For every game $v$, we have $\mathcal{C}(v) \subset \mathcal{W}(v)$ (see [7, 19]). It is well known that a game $v$ is convex if and only if $\mathcal{C}(v) = \mathcal{W}(v)$ (see [8, 9, 18, 19] and Remark 4.1).

## 3 Fuzzy Games and Fuzzy Extensions

### 3.1 Fuzzy Cores and Superdifferentials

Following [1, 2], we introduce the notions of (cooperative) fuzzy games. A vector $s \in [0,1]^n$ is called a *fuzzy coalition* and it is interpreted as a profile of the "rate of participation" of the players. The fuzzy coalition $\lambda \mathbb{1}_N \in [0,1]^n$ with $\lambda \in [0,1]$ describes that the commitment of each player is the same level $\lambda$. A function $\Gamma : [0,1]^n \to \mathbb{R}$ with $\Gamma(\mathbf{0}) = 0$ is called a *fuzzy game*, denoted alternatively by $(N, \Gamma)$. A vector $x \in \mathbb{R}^n$ is a *feasible payoff* of a fuzzy game $\Gamma$ if $x(N) \leq \Gamma(\mathbb{1}_N)$. The *fuzzy core* of $\Gamma$ is defined by the set

$$\mathcal{C}^f(\Gamma) = \{x \in \mathbb{R}^n \mid x(N) = \Gamma(\mathbb{1}_N) \text{ and } \langle s, x \rangle \geq \Gamma(s)\ \forall s \in [0,1]^n\}.$$

A fuzzy game $\Gamma$ is *balanced* if $\mathcal{C}^f(\Gamma)$ is nonempty.

Throughout this paper, we restrict our focus to the fuzzy games satisfying positive homogeneity: $\Gamma(\lambda s) = \lambda \Gamma(s)$ for every $s \in [0,1]^n$ and $\lambda \geq 0$ with $\lambda s \in [0,1]^n$. Then $\Gamma$ is extended naturally to the entire domain $\mathbb{R}^n$ by

$$s \mapsto \begin{cases} \|s\| \Gamma\left(\dfrac{s}{\|s\|}\right) & \text{if } s \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}, \\ -\infty & \text{if } s \notin \mathbb{R}^n_+ \end{cases} \tag{3.1}$$

preserving positive homogeneity on $\mathbb{R}^n$, i.e.,

(*Positive homogeneity*): $\Gamma(\lambda s) = \lambda\Gamma(s)$ for every $s \in \mathbb{R}^n$ and $\lambda \geq 0$.

For this extension we retain the same notation $\Gamma$. The advantage of positively homogeneous fuzzy games is exemplified with the following useful observation due to [1, Proposition 11.1.3.2].

**Theorem 3.1.** *If $(N, \Gamma)$ is a positively homogeneous fuzzy game, then $\mathcal{C}^f(\Gamma) = \partial\Gamma(\lambda \mathbb{1}_N)$ for every $\lambda \geq 0$.*

Hence, the investigation of the properties of fuzzy cores is reduced to the superdifferential calculus of positively homogeneous fuzzy games with the aid of convex analysis. In particular, the fuzzy core of a positively homogeneous concave fuzzy game is nonempty, compact and convex (see [1, Proposition 11.1.3]). For a fuzzy game $\Gamma$, the directional derivative $\Gamma'(s; \mathbb{1}_{\{i\}})$ is called the *limiting marginal contribution* of player $i \in N$ at fuzzy coalition $s \in [0,1]^n$ and it is well defined for every $s \in \mathbb{R}^n_+$ whenever $\Gamma$ is positively homogeneous and concave on $[0,1]^n$ because it is extended to $\mathbb{R}^n_+$ by (3.1) preserving these properties.

A *fuzzy extension* of a game $(N, v)$ is a fuzzy game $\Gamma^v : [0,1]^n \to \mathbb{R}$ such that $\Gamma^v(\mathbb{1}_S) = v(S)$ for every $S \in 2^N$. There are numerous fuzzy extensions of a given game. On the contrary, the restriction of a fuzzy game $\Gamma : [0,1]^n \to \mathbb{R}$ to characteristic functions yields a *crisp game* $v_\Gamma : 2^N \to \mathbb{R}$ defined by $v_\Gamma(S) = \Gamma(\mathbb{1}_S)$ for $S \in 2^N$. There are innumerable fuzzy games that yield a same crisp game. It is evident that $\mathcal{C}^f(\Gamma^v) \subset \mathcal{C}(v)$ and $\mathcal{C}^f(\Gamma) \subset \mathcal{C}(v_\Gamma)$. A natural question arises here. Under what condition on $v$, $\Gamma^v$ and $\Gamma$, does the coincidence $\mathcal{C}(v) = \mathcal{C}^f(\Gamma^v)$ and $\mathcal{C}^f(\Gamma) = \mathcal{C}(v_\Gamma)$ happens? It is a main concern of this paper.

A partial answer to the coincidence $\mathcal{C}^f(\Gamma) = \mathcal{C}(v_\Gamma)$ was proposed in [3] under the supermodularity and separate convexity of $\Gamma$. Supermodularity implies the convexity of the crisp game $v_\Gamma$, which is satisfied for many applications. However, as we shall see below, most useful fuzzy extensions of games do not fulfill separate convexity, but satisfy concavity. In particular, the convexity of games is equivalent to the concavity of the Lovász extensions of those. To deal with Lovász extensions, the most important class of fuzzy extensions, more suitably, an alternative hypothesis on fuzzy games should be pursued. We touch on this problem in Subsection 4.1.

### 3.2 Exact Fuzzy Extensions

In view of the definition of an exact game $v$, the support function $\Gamma : \mathbb{R}^n_+ \to \mathbb{R}$ of the core $\mathcal{C}(v)$ defined by $\Gamma(s) = \min_{x \in \mathcal{C}(v)} \langle s, x \rangle$ for $s \in \mathbb{R}^n_+$ gives rise to a possible fuzzy extension of $v$. As $\Gamma$ is a pointwise minimum of family of continuous linear function $s \mapsto \langle s, x \rangle$ over $x$ in $\mathcal{C}(v)$, it is upper semicontinuous, superadditive, positively homogeneous and translation invariant. This observation suggests that it is quite natural to introduce the functions possessing these properties as suitable extensions of exact games (see [12, 15]).

**Definition 3.1.** *A positively homogeneous fuzzy game $\Gamma : \mathbb{R}^n_+ \to \mathbb{R}$ is exact if the following conditions are satisfied.*

*(Upper semicontinuity): $\Gamma$ is upper semicontinuous.*
*(Superadditivity): $\Gamma(s+t) \geq \Gamma(s) + \Gamma(t)$ for every $s, t \in \mathbb{R}^n_+$.*
*(Translation invariance): $\Gamma(s + \lambda \mathbb{1}_N) = \Gamma(s) + \lambda \Gamma(\mathbb{1}_N)$ for every $s \in \mathbb{R}^n_+$ and $\lambda \geq 0$.*

The following characterization of exact games is due to [15, Theorem 2.2].

**Theorem 3.2.** *Let $(N, v)$ be a game. Then the following conditions are equivalent.*

(i) *$v$ is exact.*
(ii) *$v$ has an exact fuzzy extension $\Gamma^v : \mathbb{R}^n_+ \to \mathbb{R}$ such that*

$$\mathcal{C}(v) = \partial \Gamma^v(\lambda \mathbb{1}_N) \quad \text{for every } \lambda \geq 0.$$

(iii) *$v$ has an exact fuzzy extension $\Gamma^v : \mathbb{R}^n_+ \to \mathbb{R}$ such that*

$$\Gamma^v(s) = (\Gamma^v)'(\lambda \mathbb{1}_N; s) = \min_{x \in \mathcal{C}(v)} \langle s, x \rangle = \langle s, y \rangle$$

*for every $s \in \mathbb{R}^n_+$, $y \in \partial \Gamma^v(s)$ and $\lambda \geq 0$.*

The above theorem implies that for an exact game $v$ and its exact fuzzy extension $\Gamma^v$, the *limiting marginal contribution* of player $i \in N$ at the diagonal $\lambda \mathbb{1}_N$ is given by

$$(\Gamma^v)'(\lambda \mathbb{1}_N; \mathbb{1}_{\{i\}}) = \min_{x \in \mathcal{C}(v)} x_i = v(\{i\}) \quad \text{for every } \lambda \geq 0. \tag{3.2}$$

[16] defined the fuzzy extension $\Gamma^v_{\mathrm{S}} : \mathbb{R}^n_+ \to \mathbb{R}$ of an exact game $v$ by

$$\Gamma^v_{\mathrm{S}}(s) = \max \left\{ \sum_{S \in 2^N} \alpha_S v(S) - \alpha_N v(N) \,\middle|\, \begin{array}{l} \sum_{S \in 2^N} \alpha_S \mathbb{1}_S - \alpha_N \mathbb{1}_N = s \\ \alpha_S \geq 0 \ \forall S \in 2^N \end{array} \right\}.$$

This leads to another characterization of exact games.

**Theorem 3.3.** *A game $(N, v)$ is exact if and only if $\Gamma^v_{\mathrm{S}} : \mathbb{R}^n_+ \to \mathbb{R}$ is a fuzzy extension of $v$ such that*

$$\Gamma^v_{\mathrm{S}}(s) = \min_{x \in \mathcal{C}(v)} \langle s, x \rangle \quad \text{for every } s \in \mathbb{R}^n_+.$$

It follows from Theorem 3.3 that the "consistency" of the exact game and the fuzzy extension holds in the sense that

$$v \overset{\text{extension}}{\longrightarrow} \Gamma^v_{\mathrm{S}} \overset{\text{restriction}}{\longrightarrow} v_{\Gamma^v_{\mathrm{S}}} = v.$$

Moreover, the coincidence of the core holds for every exact game $v$ and the fuzzy extension $\Gamma^v_{\mathrm{S}}$.

**Corollary 3.1.** *If $v$ is exact, then $\mathcal{C}(v) = \mathcal{C}^f(\Gamma^v_{\mathrm{S}}) = \partial \Gamma^v_{\mathrm{S}}(\lambda \mathbb{1}_N)$ for every $\lambda \geq 0$.*

# 4 Coincidence of Weber Sets

## 4.1 Lovász Extensions

For an arbitrarily given vector $s = (s_1, \ldots, s_n) \in \mathbb{R}^n$, let $\pi \in \Pi_N$ be a permutation with $s_{\pi(1)} \geq \cdots \geq s_{\pi(n)}$. Define $S_i = \{j \in N \mid s_j \geq s_{\pi(i)}\}$ for each $i \in N$. A function $\Gamma_{\mathrm{L}}^v : \mathbb{R}^n \to \mathbb{R}$ is the *Lovász extension* of a game $v : 2^N \to \mathbb{R}$ if it is of the form[1]

$$\Gamma_{\mathrm{L}}^v(s) = \sum_{i=1}^{n-1} (s_{\pi(i)} - s_{\pi(i+1)}) v(S_i) + s_{\pi(n)} v(S_n).$$

It follows from the definition that $\Gamma_{\mathrm{L}}^v(\mathbb{1}_S) = v(S)$ for every $S \in 2^N$ and $\Gamma_{\mathrm{L}}^v$ satisfies positive homogeneity, Lipschitz continuity with rank $|v(N)|$ and translation invariance. Unlike the exact fuzzy extensions investigated in the previous section, every game is extendable via its Lovász extensions as a fuzzy extension.

It is known that $\partial\Gamma_{\mathrm{L}}^v(\mathbb{1}_N) = \mathcal{C}(v)$ (see [6, Proposition 3]). On the other hand, $\partial\Gamma_{\mathrm{L}}^v(\lambda\mathbb{1}_N) = \mathcal{C}^f(\Gamma_{\mathrm{L}}^v)$ for every $\lambda \geq 0$ by Theorem 3.1. Therefore, we obtain the coincidence of the core for any game.

**Theorem 4.1.** *$\mathcal{C}(v) = \mathcal{C}^f(\Gamma_{\mathrm{L}}^v) = \partial\Gamma_{\mathrm{L}}^v(\lambda\mathbb{1}_N)$ for every $\lambda \geq 0$.*

A game is convex if and only if its Lovász extension is superadditive (see [13, Theorem 4.6]). Thus, the Lovász extension of a game is an exact fuzzy extension if and only if the game is convex (see [10, Proposition 4.1]), and if and only if it is the support functional of the core, i.e.,

$$\Gamma_{\mathrm{L}}^v(s) = \min_{x \in \mathcal{C}(v)} \langle s, x \rangle \quad \text{for every } s \in \mathbb{R}^n. \tag{4.1}$$

(See [13, Theorem 4.7].)

To characterize Lovász extensions further, we introduce another condition on fuzzy games along the lines of [17]. Vectors $s$ and $t$ in $\mathbb{R}^n$ are *comonotonic* if $(s_i - s_j)(t_i - t_j) \geq 0$ for every $i, j \in N$.

(*Comonotonic additivity*): $\Gamma(s+t) = \Gamma(s) + \Gamma(t)$ for every pair of comonotonic $s, t$ in $\mathbb{R}^n$.

Recall that $v_\Gamma$ is a crisp game obtained from the restriction of a fuzzy game $\Gamma$ to characteristic functions. A fuzzy game $\Gamma$ is positively homogeneous and comonotone additive if and only if it is represented as a Lovász extension of $v_\Gamma$, that is, $\Gamma = \Gamma_{\mathrm{L}}^{v_\Gamma}$ (see [17, Proposition 1]). This result implies the "consistency"

[1] It is easy to verify that Lovász extension $\Gamma_{\mathrm{L}}^v$ coincides with the *Choquet integral* of $v$:

$$\Gamma_{\mathrm{L}}^v(s) = \int_0^{+\infty} v(\{i \in N \mid s_i \geq q\}) dq + \int_{-\infty}^0 [v(\{i \in N \mid s_i \geq q\}) - v(N)] dq$$

for every $s \in \mathbb{R}^n$.

of a fuzzy game with positive homogeneity and comonotone additivity and the Lovász extension of its restriction in the sense that

$$\Gamma \overset{\text{restriction}}{\longrightarrow} v_\Gamma \overset{\text{extension}}{\longrightarrow} \Gamma_{\mathrm{L}}^{v_\Gamma} = \Gamma. \tag{4.2}$$

Furthermore, as $\partial\Gamma_{\mathrm{L}}^{v_\Gamma}(\lambda\mathbb{1}_N) = \mathcal{C}(v_\Gamma) = \mathcal{C}^f(\Gamma_{\mathrm{L}}^{v_\Gamma}) = \mathcal{C}^f(\Gamma)$ by Theorem 4.1, the following coincidence is true.

**Theorem 4.2.** *If $(N, \Gamma)$ is a positively homogeneous, comonotone additive fuzzy game, then $\mathcal{C}^f(\Gamma) = \mathcal{C}(v_\Gamma) = \partial\Gamma(\lambda\mathbb{1}_N)$ for every $\lambda \geq 0$.*

For another characterization of Lovász extension of monotone games in terms of fuzzy extensions satisfying monotonicity, linearity with respect to games, positive homogeneity and translation invariance, see [11].

### 4.2 Fuzzy Weber Sets

Let $\Gamma : \mathbb{R}^n \to \mathbb{R}$ be a positively homogeneous, locally Lipschitz fuzzy game. If $\Gamma$ has the directional derivative $\Gamma'(s; \mathbb{1}_{\{i\}})$, then it is called the *limiting marginal contribution* of player $i \in N$ at fuzzy coalition $s \in [0,1]^n$. Denote by $a^s(\Gamma) = (\Gamma'(s; \mathbb{1}_{\{1\}}), \ldots, \Gamma'(s; \mathbb{1}_{\{n\}})) \in \mathbb{R}^n$ the vector of limiting marginal contributions at $s \in [0,1]^n$.

Let $D = \{s \in \mathbb{R}^n \mid \exists\, i, j \in N : s_i = s_j\}$. For each distinct $i, j \in N$ there exists no $\pi \in \Pi_N$ such that $P_i^\pi = P_j^\pi = \emptyset$. Hence, there exists no $\pi \in \Pi_N$ such that $a_i^\pi(v) = v(\{i\})$ and $a_j^\pi(v) = v(\{j\})$. This observation suggests that the vector of limiting marginal contributions $a^s(\Gamma)$ at any $s \in D$ should be excluded from the definition of the fuzzy Weber sets to keep the consistency with the definition of the Weber sets.

**Definition 4.1.** *The fuzzy Weber set of $(N, \Gamma)$ is defined by the set*

$$\mathcal{W}^f(\Gamma) = \mathrm{co}\{a^s(\Gamma) \in \mathbb{R}^n \mid s \in [0,1]^n \setminus D\}.$$

For fuzzy Weber sets to be a reasonable solution concept, any payoff vector in $\mathcal{W}^f(\Gamma)$ needs to be feasible. This is indeed true for a fuzzy game $\Gamma$ such that the directional derivative $\Gamma'(s; h)$ exists for every $s \in [0,1]^n$ and $h \in \mathbb{R}^n$. To this end, it suffices to show that $\sum_{i \in N} \Gamma'(s; \mathbb{1}_{\{i\}})) \leq \Gamma(\mathbb{1}_N)$ for every $s \in [0,1]^n$, but this inequality trivially follows from the fact that $h \mapsto \Gamma'(s; h)$ is positively homogeneous and concave (see [14, Theorem 23.1]).

We are ready to provide the main result of this paper, the coincidence of the Weber and fuzzy Weber sets for "every" game. This result is more significant than the core coincidence because Theorem 4.1 is meaningless when the core is empty, but the coincidence of Weber sets is always valid without any assumption on games.

**Proposition 4.1.** *$\mathcal{W}(v) = \mathcal{W}^f(\Gamma_{\mathrm{L}}^v) = \partial_{\mathrm{C}}\Gamma_{\mathrm{L}}^v(\lambda\mathbb{1}_N)$ for every $\lambda \geq 0$.*

**Corollary 4.1.** *If $(N, \Gamma)$ is a positively homogeneous, comonotone additive fuzzy game, then $\mathcal{W}^f(\Gamma) = \mathcal{W}(v_\Gamma) = \partial_{\mathrm{C}}\Gamma(\lambda\mathbb{1}_N)$ for every $\lambda \geq 0$.*

### 4.3 Regularity and Convexity

Although $\Gamma_{\mathrm{L}}^v$ always possesses the Clarke directional derivative $(\Gamma_{\mathrm{L}}^v)^\circ(s;\mathbb{1}_{\{i\}})$ for every $s \in [0,1]^n$ in view of the Lipschitz continuity, there is no guarantee that it coincides with the limiting marginal contribution $(\Gamma_{\mathrm{L}}^v)'(s;\mathbb{1}_{\{i\}})$. Since the inequality $(\Gamma_{\mathrm{L}}^v)^\circ(s;\mathbb{1}_{\{i\}}) \leq (\Gamma_{\mathrm{L}}^v)'(s;\mathbb{1}_{\{i\}})$ is true by definition, the Clarke directional derivative underestimates the limiting marginal contribution of each player. This discrepancy disappears whenever $\Gamma_{\mathrm{L}}^v$ is regular and thereby it is reasonable to introduce the notion of regularity of games, which is a mild condition on the "smoothness" of $\Gamma_{\mathrm{L}}^v$.

**Definition 4.2.** *A game $(N,v)$ is regular if its Lovász extension $\Gamma_{\mathrm{L}}^v$ is regular at the origin.*

Here, the regularity of $v$ means that $(\Gamma_{\mathrm{L}}^v)^\circ(\mathbf{0};h) = (\Gamma_{\mathrm{L}}^v)'(\mathbf{0};h)$ for every $h \in \mathbb{R}^n$.

The third remarkable property of the Lovász extension is that the regularity of $\Gamma_{\mathrm{L}}^v$ is equivalent to the convexity of $v$. Specifically, we can derive an additional characterization of convex games as follows.

**Theorem 4.3.** *Let $(N,v)$ be a game. Then the following conditions are equivalent.*

(i) *$v$ is convex.*
(ii) *$v$ is regular.*
(iii) *$\Gamma_{\mathrm{L}}^v$ is concave.*
(iv) *$\Gamma_{\mathrm{L}}^v(s) = \min_{x\in\mathcal{C}(v)} \langle s,x\rangle = \min_{x\in\mathcal{W}(v)} \langle s,x\rangle$ for every $s \in \mathbb{R}^n$.*
(v) *$\mathcal{C}(v) = \mathcal{C}^f(\Gamma_{\mathrm{L}}^v) = \mathcal{W}^f(\Gamma_{\mathrm{L}}^v) = \mathcal{W}(v)$.*

*Remark 4.1.* The equivalence of condition (i) and the coincidence $\mathcal{C}(v) = \mathcal{W}(v)$ in Theorem 4.3 is well known. [8, Theorem (22)] and [18, Theorems 3 and 5] demonstrated independently the inclusion $\mathcal{W}(v) \subset \mathcal{C}(v)$ whenever $v$ is convex. The converse inclusion $\mathcal{C}(v) \subset \mathcal{W}(v)$ for convex game $v$ was established by [9, Corollary 1]. It was proven by [19, Theorem 14] that the inclusion $\mathcal{C}(v) \subset \mathcal{W}(v)$ holds for every game $v$. The equivalence (i) $\Leftrightarrow$ (iii) is attributed to [4, p. 289]. Regularity in condition (ii) and its relevance to (fuzzy) Weber sets are new in the literature.

### 4.4 Regularity and Supermodularity

Let $\vee$ and $\wedge$ be the lattice operations in $\mathbb{R}^n$ defined by

$$s \vee t = (\max\{s_1,t_1\},\ldots,\max\{s_n,t_n\})$$

and

$$s \wedge t = (\min\{s_1,t_1\},\ldots,\min\{s_n,t_n\}).$$

(*Supermodularity*): $\Gamma(s) + \Gamma(t) \leq \Gamma(s\vee t) + \Gamma(s\wedge t)$ for every $s,t \in \mathbb{R}^n$.

**Theorem 4.4.** *Let $(N, \Gamma)$ be a positively homogeneous fuzzy game. Then the following conditions are equivalent.*

(i) *$\Gamma$ is supermodular and translation invariant.*
(ii) *$\Gamma$ is concave and $\mathcal{C}^f(\Gamma) = \mathcal{C}(v_\Gamma) = \mathcal{W}(v_\Gamma) = \mathcal{W}^f(\Gamma)$.*
(iii) *$\Gamma$ is regular and $\mathcal{C}^f(\Gamma) = \mathcal{C}(v_\Gamma) = \mathcal{W}(v_\Gamma) = \mathcal{W}^f(\Gamma)$.*

## References


[1] Aubin, J.-P.: Mathematical Methods of Game and Economic Theory. North-Holland, Amsterdam (1979)
[2] Aubin, J.-P.: Cooperative fuzzy games. Math. Oper. Res. 6, 1–13 (1981)
[3] Branzei, R., Dimitrov, D., Tijs, S.: Convex fuzzy games and participation monotonic allocation schemes. Fuzzy Sets and Systems 139, 267–281 (2003)
[4] Choquet, G.: Theory of capacities. Ann. Inst. Fourier (Grenoble) 5, 131–295 (1954)
[5] Clarke, F.H.: Optimization and Nonsmooth Analysis. John Wiley & Sons, New York (1983)
[6] Danilov, V., Koshevoy, G.A.: Cores of cooperative games, superdifferentials of functions, and the Minkowski difference of sets. J. Math. Anal. Appl. 247, 1–14 (2000)
[7] Derks, J.J.M.: A short proof of the inclusion of the core in the Weber set. Internat. J. Game Theory 21, 149–150 (1992)
[8] Edmonds, J.: Submodular functions, matroids, and certain polyhedra. In: Guy, R., et al. (eds.) Combinatorial Structures and their Applications, pp. 69–87. Gordon & Breach, New York (1970)
[9] Ichiishi, T.: Super-modularity: Applications to convex games and to the greedy algorithm for LP. J. Econom. Theory 25, 283–286 (1981)
[10] Lovász, L.: Submodular functions and convexity. In: Bachem, A., et al. (eds.) Mathematical Programming–The State of the Art, pp. 235–257. Springer, Berlin (1983)
[11] Marichal, J.-L.: An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. IEEE Trans. Fuzzy Systems 8, 800–807 (2000)
[12] Maaß, S.: Exact functionals and their core. Statist. Papers 43, 75–93 (2002)
[13] Marinacci, M., Montrucchio, L.: Introduction to the mathematics of ambiguity. In: Gilboa, I. (ed.) Uncertainty in Economic Theory, pp. 46–107. Routledge, New York (2004)
[14] Rockafellar, R.T.: Convex Analysis. Princeton Univ. Press, Princeton (1970)
[15] Sagara, N.: A probabilistic characterization of exact games on $\sigma$-algebras. Fuzzy Sets and Systems 216, 34–51 (2013)
[16] Schmeidler, D.: Cores of exact games, I. J. Math. Anal. Appl. 40, 214–225 (1972)
[17] Schmeidler, D.: Integral representation without additivity. Proc. Amer. Math. Soc. 97, 255–261 (1986)
[18] Shapley, L.S.: Cores of convex games. Internat. J. Game Theory 1, 11–26 (1971)
[19] Weber, R.: Probabilistic values for games. In: Roth, A.E. (ed.) The Shapley Value: Essays in Honor of Lloyd S. Shapley, pp. 101–119. Cambridge University Press, Cambridge (1988)

# Set Function Representations of Alternatives' Relative Features

Eiichiro Takahagi

Senshu University, Kawasaki 2148580, Japan
takahagi@isc.senshu-u.ac.jp

**Abstract.** In a weighted sum model such as the Analytic Hierarchy Process, a set function value is constructed from the weights of the model. In this paper, we use relative individual scores to propose a set function that shows the features of an alternative. The set function value for the alternative is calculated by averaging the values of the set function representation of the weights generated when the alternative has the highest comprehensive score. By interpreting the functions, we can understand the features of an alternative. We discuss the properties of the set functions, and extend to Choquet integral models.

**Keywords:** alternatives, multi-criteria decision-making, set functions, fuzzy measures, Choquet integral, Shapley value, Möbius transformation.

## 1 Introduction

In multi-criteria decision-making models, such as the Analytic Hierarchy Process [3] and Choquet integral models [2][4], it is useful to display the features of an alternative. The figures are determined from the relative evaluation scores among alternatives. There are some works to explain of a decision such as [6], [8], and [9]. Especially in [9], explanations are generated by analyzing whether the decision would be taken by changing the values of weights. In this paper, by changing the values of weights, for an alternative, we show which set of criteria is important diffrent conditions on the comparison of weights and the figure of the degree. The aim of this paper is to define the figures using set functions, and analyze the set function properties.

**Table 1.** Example 1 (Two Criteria and Four Alternatives Model)

| Alternatives | Criterion 1 (C1) | Criterion 2 (C2) |
|---|---|---|
| A1 | 0.35 | 0.15 |
| A2 | 0.05 | 0.50 |
| A3 | 0.24 | 0.34 |
| A4 | 0.36 | 0.01 |

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 246–255, 2014.

Table 1 is an example of a two-criteria (C1 and C2) and four-alternatives (A1, ..., A4) multi-criteria decision-making model. In the weighted sum model, if a decision maker (DM) sets greater importance on C1, then the alternative A1 is selected, because it has a large score (0.35) for C1 and a modest score (0.15) for C2. If the DM sets extremely high importance on C1, then A4 is selected, because the A4 score for C1 is a little higher than that of A1. A2 is selected if the DM sets greater importance on C2, and A3 is selected if C1 and C2 are considered to be equally important.

Using these relations among alternatives scores, we propose the set function representation $E_{X,j}(A)$ of alternative $j$. Figure 2 illustrates the result of the proposed set function.

In [10], we proposed the concept of set functions. In this paper, we propose a uniform generating method, compare this with the random generating method, and extend the concept to Choquet integral models.

## 2 Notation

Let $Y = \{1, \ldots, n\}$ be the set of criteria ($n$: number of criteria), and $m$ be the number of alternatives. $x_i^j \in [0,1], i = 1, \ldots, n, j = 1, \ldots, m)$ is the individual score of the $i$-th criterion and $j$-th alternative. The $x_i^j$ have the property of strong commensurability. $X$ is the matrix representation of $x_i^j$. In this paper, $X$ is given in advance, such as by pairwise comparisons.

$\boldsymbol{w}^k = (w_1^k, \ldots, w_n^k), w_i^k \in [0,1], \sum_i w_i^k = 1$ gives the weight of each criterion. As we calculate various cases by varying $\boldsymbol{w}$, we use the superscript $k$.

The comprehensive score $y_j^k$ is calculated as the weighted sum

$$y_j^k = \sum_{i=1}^{n} (w_i^k x_i^j). \tag{1}$$

A fuzzy measure $\mu$ is defined as

$$\mu : 2^Y \to [0,1] \text{ where } \mu(\emptyset) = 0 \text{ and } \mu(A) \geq \mu(B) \text{ if } A \supseteq B. \tag{2}$$

The comprehensive score $y_j^k$ calculated by Choquet integral is

$$y_j^k = \sum_{i=1}^{n} [x_{\sigma(i)}^j - x_{\sigma(i+1)}^j] \mu(\{\sigma(1), \ldots, \sigma(i)\}) \tag{3}$$

where $\sigma(i)$ is the permutation on $Y$, that is, $x_{\sigma(1)}^j \geq \ldots \geq x_{\sigma(n)}^j$, $Y = \{\sigma(1), \ldots, \sigma(n)\}$, $\sigma(n+1) = n+1$, and $x_{n+1}^j = 0$. The Shapley value of $\mu$ [1][5] is

$$\phi_i(\mu) \equiv \sum_{S \subseteq Y} \gamma_n(S)[\mu(S) - \mu(S \setminus \{i\})] \text{ where } \gamma_n(S) = \frac{(n - |S|)!(|S| - 1)!}{n!} \tag{4}$$

and the Möbius transformation [7] is

$$\nu(A) \equiv \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B) \text{ and } \mu(A) = \sum_{B \subseteq A} \nu(B), \forall A \in 2^Y. \tag{5}$$

## 3 Set Function Representations (Weighted Sum Cases)

### 3.1 Set Functions

$E_{X,j}(A)$ shows the degree to which alternative $j$ is the best alternative when criteria in $A$ have the large weights. If $E_{X,j}(A)$ has a high score, we interpret that alternative $j$ is usually selected when $w_i, \forall i \in A$ are large values. For some $X$, $E_{X,j}(A)$ is defined as

$$E_{X,j} : 2^Y \to \boldsymbol{R}^+, \;\; E_{X,j}(\emptyset) = 0, j = 1, \ldots, m. \tag{6}$$

When $n = 2$, $Y = \{1, 2\}$, $E_{X,j}$ is constituted of $E_{X,j}(\emptyset)$, $E_{X,j}(\{1\})$, $E_{X,j}(\{2\})$, and $E_{X,j}(\{1, 2\})$. $E_{X,j}(\{1\})$ shows the degree to which alternative $j$ has the highest comprehensive score when only $w_1$ is assigned a greater score. $E_{X,j}(\{1, 2\})$ shows the degree to which alternative $j$ has the highest comprehensive score when both $w_1$ and $w_2$ are assigned greater values, that is, cases close to $w_1 = w_2$.

### 3.2 Set Function for $\boldsymbol{w^k}$

$E^k(A)$ shows the reconstituted weight of set $A$ and is calculated by rank depended method such as the Choquet integrals. For example, if $n = 3$ and $w_2^k \geq w_3^k \geq w_1^k$, then $E^k(\{2\}) = w_2^k - w_3^k$, $E^k(\{2, 3\}) = w_3^k - w_1^k$, and $E^k(\{1, 2, 3\}) = w_1^k$.

**Definition 1 (Set Function for $\boldsymbol{w}$).** *For some $\boldsymbol{w}^k$, the set function $E^k$ is*

$$E^k(\{\sigma(1), \ldots, \sigma(i)\}) = [w_{\sigma(i)}^k - w_{\sigma(i+1)}^k], \tag{7}$$

*$(i = 1, \ldots, n)$ where $\sigma(i)$ is the permutation on $Y$, that is, $w_{\sigma(1)}^k \geq \ldots \geq w_{\sigma(n)}^k$, $Y = \{\sigma(1), \ldots, \sigma(n)\}$, $\sigma(n + 1) = n + 1$, and $w_{n+1}^k = 0$. Values that are not assigned by eq. (7) are zero, that is, $E^k(A) = 0, \forall A \in (2^Y \backslash \bigcup_{i=1}^n \{\{\sigma(1), \ldots, \sigma(i)\}\})$.*

For $A \in \{\{\sigma(1)\}, \{\sigma(1), \sigma(2)\}, \ldots, \{\sigma(1), \ldots, \sigma(n)\}\}$, the $E^k(A)$ can be active, $E^k(A) > 0$. $(\{\sigma(1)\}, \{\sigma(1), \sigma(2)\}, \ldots, \{\sigma(1), \ldots, \sigma(n)\}))$ is the maximum chain of $Y$.

**Theorem 1.** *For any $\boldsymbol{w}^k$ and $i$,*

$$\sum_{B \ni i} E^k(B) = w_i^k, \tag{8}$$

$$\sum_{A \in 2^Y} [\,|\, A \,|\, E^k(A)] = 1. \tag{9}$$

*Proof. Eq. (8) is trivial.*

$$\sum_{A \in 2^Y} [\,|\, A \,|\, E^k(A)] = \sum_{i=1}^{n} [\,|\, \{\sigma(1), \ldots, \sigma(i)\} \,|\, E^k(\{\sigma(1), \ldots, \sigma(i)\})]$$

$$= \sum_{i=1}^{n} [i(w_{\sigma(i)}^k - w_{\sigma(i+1)}^k)] = \sum_{i=1}^{n} w_{\sigma(i)}^k = 1.$$

For example, if $\boldsymbol{w}^k = (0.7, 0.3)$, then $E^k(\{1\}) = 0.7 - 0.3 = 0.4$, $E^k(\{2\}) = 0$, $E^k(\{1,2\}) = 0.3$, and $E^k(\{1\}) + E^k(\{2\}) + 2E^k(\{1,2\}) = 1$.

### 3.3 Set Functions for the Alternative $j$

For weights $\boldsymbol{w}^k$, if the alternative $j$ is selected, the set function $E^k(A), \forall A$ belongs to alternative $j$. By generating multiple $\boldsymbol{w}^k, k = 1, \ldots, K$, we calculate $E^*_{X,j}$. $q^k_j$ is a flag for when alternative $j$ is selected for $\boldsymbol{w}^k$. If two or more alternatives are selected for $\boldsymbol{w}^k$, $q^k_j$ is assigned on a pro-rata basis, that is,

$$H^k = \{j \mid y^k_j \geq y^k_l, \forall l = 1, \ldots, m\}, \tag{10}$$

$$q^k_j = \begin{cases} 1/\mid H^k \mid & \text{if } j \in H^k \\ 0 & \text{otherwise} \end{cases}, \quad Q_j = \sum_{k=1}^{K} q^k_j. \tag{11}$$

Obviously, $\sum_j q^k_j = 1 \forall k$ and $\sum_{j=1}^{m} Q_j = K$.

As the selection of the best alternative is depended on weights and individual scores, the set function $E^*_{X,j}$ that shows the relative features of alternative $j$ is defined as the average value of $E^k(A)$ when alternative $j$ is selected. In this paper, we calculate the set function using simulations.

**Definition 2 ($E^*_{X,j}$).** *$E^+_{X,j}$ and $E^*_{X,j}$ are defined as*

$$E^+_{X,j}(A) = \sum_{k=1}^{K} [q^k_j E^k(A)], \tag{12}$$

$$E^*_{X,j}(A) = E^+_{X,j}(A)/Q_j, \forall A \in 2^X. \tag{13}$$

Let $d > 0$ be the simulation number (positive integer). In this method, we use $w_i \in \{0, 1/d, 2/d, \ldots, d/d\}$.

**Definition 3 (Uniform Generating Method).** *$\Omega^U$ is the set of all weights $\boldsymbol{w}^k$ for the simulation.*

$$\Omega^U = \{\boldsymbol{w} \mid w_i \in \{0, 1/d, 2/d, \ldots, d/d\} \text{ where } \sum w_i = 1 \ \} \tag{14}$$

We number the elements of $\Omega^U$, that is, $\Omega^U = \{\boldsymbol{w}^1, \ldots, \boldsymbol{w}^K\}$.

**Definition 4 (Random Generating Method).** *$w'^k_i$ is assigned by uniform random numbers in $(0, 1)$. $w^k_i$ is given by $w^k_i = \frac{w'^k_i}{\sum_i w'^k_i} \forall k$, and $\Omega^R = \{\boldsymbol{w}^1, \ldots, \boldsymbol{w}^K\}$.*

The average and standard deviation of $A \in 2^Y$ are
$E^M_X(A) \equiv [\sum_{k=1}^{K} E^k(A)]/K$ and $E^{SD}_X(A) \equiv [\sum_{k=1}^{K} (E^k(A) - E^M_X(A))^2/K]^{0.5}$.

Figure 1 illustrates the calculation process of $E^*_{X,j}(A)$ in the uniform generating method, example 1, and $d = 100$. For each $w_1$ and $w_2 = 1 - w_1$, $E^k(A)$ are calculated from eq. (7). For $w_1 \in [0, 0.457]$, A2 is selected, therefore $E^*_{X,2}(\{1\})$ is the average value of $E^k(\{1\})$ for $w_1 \in [0, 0.457]$, that is, 0.567, $E^*_{X,2}(\{2\}) = 0$, and $E^*_{X,2}(\{1,2\}) = 0.2165$.

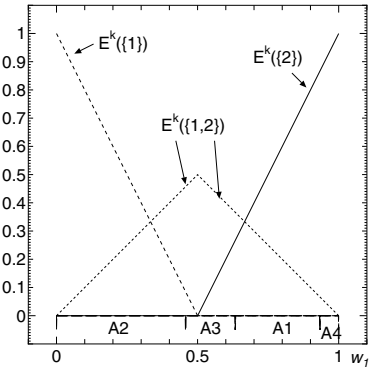



**Fig. 1.** Calculation Process

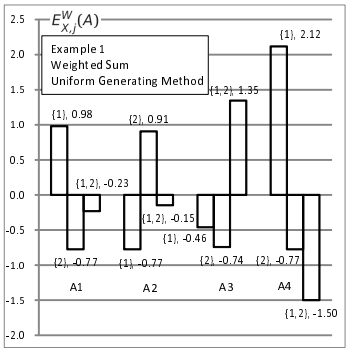



**Fig. 2.** $E^W_{X,j}$ of Example 1

### 3.4 Interpretation and Properties of the Set Functions

**Theorem 2.** *For any $j$ where $Q_j > 0$,*

$$\sum_{A \in 2^Y} [\mid A \mid E^*_{X,j}(A)] = 1 \tag{15}$$

*Proof.*

$$\sum_{A \in 2^Y} \mid A \mid E^*_{X,j}(A) = \sum_{A \in 2^Y} \mid A \mid E^+_{X,j}(A)/Q_j = \sum_{A \in 2^Y} \mid A \mid \sum_{k=1}^{K} (q_j^k E^k(A)) / \sum_{k=1}^{K} q_j^k$$

$$= \sum_{k=1}^{K} [q_j^k \sum_{A \in 2^Y} (\mid A \mid (E^k(A)))] / \sum_{k=1}^{K} q_j^k = \sum_{k=1}^{K} q_j^k / \sum_{k=1}^{K} q_j^k = 1$$

From the theorem, $E^*_{X,j}(A)$ shows the selectability of alternative $j$, because $E^*_{X,j}(A)$ is the average value when alternative $j$ is selected. As the average and standard deviation of $E^k(A)$ differ according to $\mid A \mid$, we define $E^W_{X,j}(A)$ as

**Definition 5**

$$E^W_{X,j}(A) \equiv \frac{E^*_{X,j}(A) - E^M_X(A)}{E^{SD}_X(A)}. \tag{16}$$

Figure 2 is $E^W_{X,j}(A)$ from example 1. The figure shows the relative future performance of each alternative. For example, A1 is selected when $E^k(\{1\})$ has a high score, and is not selected when $E^k(\{2\})$ has a high score. A3 is selected when $E^k(\{1,2\})$ has a high score, and is not selected when either $E^k(\{1\})$ or $E^k(\{2\})$ has a high score.

**Theorem 3.** *For any A,*

$$\sum_{j=1}^{m} Q_j E^W_{X,j}(A) = 0. \tag{17}$$

*Proof.*

$$\sum_{j=1}^{m} Q_j E^W_{X,j}(A) = \sum_{j=1}^{m} \frac{[E^+_{X,j}(A)/Q_j - E^M_X(A)]Q_j}{E^{SD}_X(A)}$$

$$= \frac{\sum_{j=1}^{m}[E^+_{X,j}(A) - E^M_X(A)Q_j]}{E^{SD}_X(A)} = \frac{[\sum_{j=1}^{m}\sum_{k=1}^{K} q^k_j E^k(A)] - K E^M_X(A)}{E^{SD}_X(A)} = 0$$

$E^W_{X,j}(A)$ can be used to compare alternatives. For this purpose, we define $E^\sharp_{X,j}(A)$ as

**Definition 6**

$$E^\sharp_{X,j}(A) \equiv E^+_{X,j}(A)/K, \qquad \forall A \in 2^Y \tag{18}$$

**Theorem 4**

$$\sum_{j=1}^{m} \sum_{A \in 2^Y} [|A|\ E^\sharp_{X,j}(A)] = 1 \tag{19}$$

*Proof*

$$\sum_{j=1}^{m} \sum_{A \in 2^Y} [|A|\ E^\sharp_{X,j}(A)] = [\sum_{j=1}^{m} \sum_{A \in 2^Y} |A|\ E^+_{X,j}(A)]/K$$

$$= \frac{1}{K} \sum_{A \in 2^Y} \sum_{j=1}^{m} [|A|\ E^+_{X,j}(A)] = \frac{1}{K} \sum_{k=1}^{K} \sum_{A \in 2^Y} [|A|\ E^k(A)] = \frac{K}{K} = 1$$

Figure 3 shows a comparison between the uniform and random generating methods. Both graphs have the same trend—if $E^k(\{1\})$ is high, A1 is usually selected, and if $E^k(\{1,2\})$ is high, A2 or A3 are usually selected.

### 3.5 Shapley Value and Möbius Transformation

As set functions are closely related to fuzzy measure theories, we use the Möbius transformation and discuss the resulting properties.

**Definition 7.** *The average importance of alternative j and criterion i is defined as*

$$S^\sharp_{X,ji} = \sum_{B \ni i} E^\sharp_{X,j}(B). \tag{20}$$

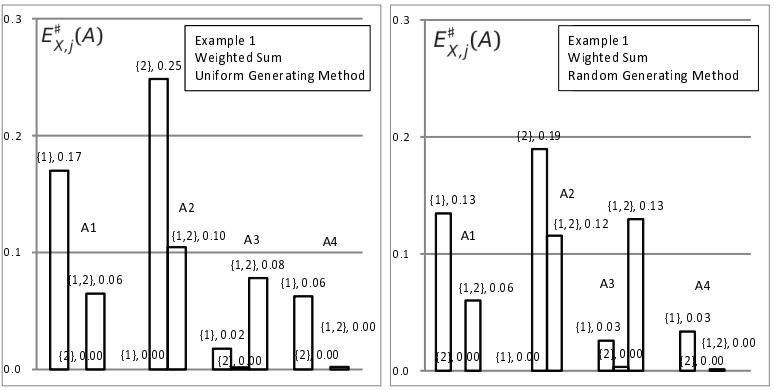



**Fig. 3.** Set Function Representation of Example 1 (Weighted Sum)

From the definition,

$$S^{\sharp}_{X,ji} = [\sum_{k=1}^{K} q_j^k w_i^k]/K. \tag{21}$$

**Definition 8 (Cumulative Set Function).** *The Cumulative Set Function* $G^{\sharp}_{X,j}$ *for* $E^{\sharp}_{X,j}$ *is defined as*

$$G^{\sharp}_{X,j}(A) = \sum_{B \subseteq A} [\mid B \mid E^{\sharp}_{X,j}(B)]. \tag{22}$$

Eq. (22) is the inverse Möbius transformation of $[\mid B \mid E^{\sharp}_{X,j}(B)]$. As $[\mid B \mid E^{\sharp}_{X,j}(B)] \geq 0 \forall B$, $G^{\sharp}_{X,j}$ are monotone and super-additive fuzzy measures. As the Shapley value of the fuzzy measure $G^{\sharp}_{X,j}$ is $\sum_{B \ni i}(\mid B \mid E^{\sharp}_{X,j}(B))/ \mid B \mid$, $S^{\sharp}_{X,ji}$ is the Shapley value of criterion $i$ when alternative $j$ is selected.

**Theorem 5.** $G^{\sharp}_{X,j}$ *and* $S^{\sharp}_{X,ji}$ *have the following properties:*

$$S^{\sharp}_{X,ji} = \frac{1}{K} \sum_{k=1}^{K} q_j^k w_i^k \tag{23}$$

$$\sum_{j=1}^{m} \sum_{i=1}^{n} S^{\sharp}_{X,ji} = 1 \tag{24}$$

$$\sum_{j=1}^{m} S^{\sharp}_{X,ji} = \frac{1}{n}, \forall i \ \ if \ \ \sum_{k=1}^{K} w_1^k = \ldots = \sum_{k=1}^{K} w_n^k \tag{25}$$

$$G^{\sharp}_{X,j}(Y) = Q_j/K \tag{26}$$

*Proof. From Theorem 1,*

$$S^{\sharp}_{X,ji} = \sum_{B \ni i} E^{\sharp}_{X,j}(B) = \frac{1}{K} \sum_{k=1}^{K} \sum_{B \ni i} q_j^k E^k(B) = \frac{1}{K} \sum_{k=1}^{K} q_j^k w_i^k.$$

$$\sum_{j=1}^{m} \sum_{i=1}^{n} S^{\sharp}_{X,ji} = \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{1}{K} \sum_{k=1}^{K} q_j^k w_i^k = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{n} w_i^k = 1.$$

*As* $\frac{1}{K} \sum_{k=1}^{K} w_i^k = \frac{1}{n}$ *from the assumption of eq. (25) and* $\sum_{i=1}^{n} \frac{1}{K} \sum_{k=1}^{K} w_i^k = 1$,

$$\sum_{j=1}^{m} S^{\sharp}_{X,ji} = \frac{1}{K} \sum_{k=1}^{K} \sum_{j=1}^{m} q_j^k w_i^k = \frac{1}{K} \sum_{k=1}^{K} w_i^k = \frac{1}{n}.$$

*From Theorem 1,*

$$G_{X,j}(Y) = \sum_{B \in 2^Y} [\mid B \mid E^{\sharp}(B)] = \frac{\sum_{B \in 2^Y} [\mid B \mid \sum_{k=1}^{k} q_j^k E^k(B)]}{K} = \frac{\sum_{k=1}^{k} q_j^k}{K} = \frac{Q_j}{K}.$$

$S^{\sharp}_{X,ji}$ is the average weight $i$ when alternative $j$ is selected and the sum of $i$ and $j$ is 1. $G_{X,j}(Y)$ shows the frequency with which alternative $j$ is selected.

## 4 Set Function Representations (Choquet Integral Cases)

In the Choquet integral cases, we divide a fuzzy measure feature into weights and interaction degrees. In this model, we use the normal fuzzy measure, that is $\mu(Y) = 1$. The weight of a fuzzy measure $\mu^k$ is the Shapley value of the fuzzy measure, and the interaction degree of a fuzzy measure $\mu^k$ is the Möbius transformation $\nu^k$ of $\mu^k$ without singletons.

**Definition 9 (Weights and Interaction Degrees).**

$$w_i^k \equiv \phi_i(\mu^k), \tag{27}$$

$$M^k(A) \equiv \begin{cases} 0 & \textit{if} \mid A \mid \leq 1 \\ \nu^k(A) & \textit{otherwise} \end{cases}. \tag{28}$$

Analogous to the weighted sum cases, by generating lots of $\mu^k$, we analyze the average set functions when the alternative is selected.

**Definition 10 (Uniform Generating Method $\mu$).** *$\Omega^F$ is the set of all fuzzy measures $\mu^k$ for the simulation,*

$$\Omega^F = \{\mu \mid \mu(A) \in \{0, 1/d, 2/d, \ldots, d/d\}, \forall A \in 2^Y \setminus \{\emptyset, Y\} \quad \textit{where } \mu \textit{ is a fuzzy measure } \}. \tag{29}$$

Let $K = |\Omega^F|$, and let $q_j^k$ and $Q_j$ be as defined in eqs. (11) , where the $y_j^k$ are calculated by eq. (3). $E^*_{X,j}$, $E^W_{X,j}$, and $E^\sharp_{X,j}$ are calculated in the same way as for the weighted sum cases.

**Definition 11.**

$$M^*_{X,j}(A) \equiv \sum_{k=1}^{k} [q_j^k M^k(A)]/Q_j, \quad M^C_{X,j}(A) \equiv \sum_{B \subseteq A} [|B| \; M^*_{X,j}(B)] \tag{30}$$

$M^*_{X,j}(A)$ shows the interaction degree among the elements of $A$. $M^C_{X,j}(A)$ shows the cumulative set function of $M^*_{X,j}(A)$. If $M^C_{X,j}(A)$ is positive, alternative $j$ is usually selected when the fuzzy measure $\mu^k$ has super-additivity among $A$.

## 5 Numerical Examples

In Table 2, $X$ denotes the model of example 2, where $n = 4$ and $m = 5$. Table 2 gives the set function representations of the weighted sum model using the uniform generating method, $d = 100$, and $K = 176,851$. As $Q_1 = 0$, A1 is not selected for any weights. As $E^\sharp_{X,3}(\{3\})$ has the highest value, A3 is usually selected when $E^k(\{3\})$ is a high value. Despite $x_2^3 = 0.4$, $E^\sharp_{X,3}(\{2\}) = 0.01$ is fairly low, because $x_2^5 = 0.44$. When $E^k(\{2\})$ has a high value, A5 is selected.

**Table 2.** Set Function Representation of Example 2 (Weighted Sum)

| | $X$ | | | | | $S^\sharp_{X,ji}$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $j$ | C1 | C2 | C3 | C4 | $Q_j$ | C1 | C2 | C3 | C4 |
| A1 | 0.18 | 0.1 | 0.2 | 0.12 | 0 | | | | |
| A2 | 0.44 | 0.01 | 0.04 | 0.03 | 10,553 | 0.04 | 0.01 | 0.01 | 0.00 |
| A3 | 0.01 | 0.4 | 0.45 | 0.05 | 53,425 | 0.05 | 0.08 | 0.14 | 0.03 |
| A4 | 0.35 | 0.05 | 0.11 | 0.5 | 75,605 | 0.13 | 0.06 | 0.08 | 0.16 |
| A5 | 0.02 | 0.44 | 0.2 | 0.3 | 37,269 | 0.03 | 0.11 | 0.03 | 0.05 |

| | $E^\sharp_{X,j}$ | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $j$ | {1} | {2} | {1,2} | {3} | {1,3} | {2,3} | {1,2,3} | {4} | {1,4} | {2,4} | {1,2,4} | {3,4} | {1,3,4} | {2,3,4} | $Y$ | |
| 1 | | | | | | | | | | | | | | | | |
| 2 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| 3 | 0.00 | 0.01 | 0.00 | 0.06 | 0.01 | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.02 | |
| 4 | 0.04 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.06 | 0.02 | 0.00 | 0.01 | 0.01 | 0.02 | 0.00 | 0.03 | |
| 5 | 0.00 | 0.05 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 | |

Table 3 gives results for the Choquet integral case using the uniform generating method, where $d = 10$ and $K = 25.683 \times 10^9$. Unlike the weighted sum model, A1 is selected when super-additivity exists between C1, C2, and C3 ($M^C_{X,1}(\{1,2,3\}) = 2.55$), or $Y$ ($M^C_{X,1}(Y) = 4.26$). A2 is usually selected when $E^k(\{1\})$ has a high value ($E^W_{X,2}(\{1\} = 3.66)$ and/or when $\{1,2\}$, $\{1,3\}$ and/or sub-additivity exists between C1, C2, and C3 ($M^C_{X,2}(\{1,2,3\}) = -3.02$).

**Table 3.** Set Function Representation of Example 2 (Choquet Integral)

| | $E^W_{X,j}$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $j$ | {1} | {2} | {1,2} | {3} | {1,3} | {2,3} | {1,2,3} | {4} | {1,4} | {2,4} | {1,2,4} | {3,4} | {1,3,4} | {2,3,4} | $Y$ |
| 1 | 0.92 | -0.39 | 0.39 | -0.04 | 1.99 | -0.32 | 0.92 | -0.40 | -0.23 | -0.33 | -0.26 | -0.31 | 0.07 | -0.43 | -0.46 |
| 2 | 3.66 | -0.41 | 0.64 | -0.40 | 0.86 | -0.33 | 0.43 | -0.41 | -0.11 | -0.33 | -0.10 | -0.33 | -0.06 | -0.43 | -0.88 |
| 3 | -0.27 | 0.13 | -0.02 | 0.54 | 0.12 | 0.43 | 0.42 | -0.36 | -0.32 | -0.18 | -0.33 | -0.03 | -0.24 | 0.16 | -0.01 |
| 4 | 0.38 | -0.37 | -0.07 | -0.29 | 0.04 | -0.32 | -0.28 | 0.35 | 0.44 | -0.11 | 0.21 | 0.04 | 0.40 | -0.29 | -0.05 |
| 5 | -0.31 | 0.38 | 0.11 | -0.13 | -0.21 | 0.01 | -0.06 | -0.10 | -0.27 | 0.34 | 0.05 | -0.02 | -0.30 | 0.24 | 0.09 |

| | | $M^C_{X,j}$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $j$ | $Q_j$ | {1,2} | {1,3} | {2,3} | {1,2,3} | {1,4} | {2,4} | {1,2,4} | {3,4} | {1,3,4} | {2,3,4} | $Y$ |
| 1 | $0.080 \times 10^9$ | 0.41 | 1.02 | -0.37 | 2.55 | -0.49 | -0.21 | 2.32 | -0.02 | 2.46 | -0.22 | 4.26 |
| 2 | $0.130 \times 10^9$ | -0.85 | -0.93 | -0.18 | -3.02 | -0.84 | 0.20 | -2.43 | 0.27 | -2.73 | 0.57 | -2.67 |
| 3 | $7.554 \times 10^9$ | 0.08 | -0.07 | 0.71 | 0.03 | -0.12 | 0.03 | 0.28 | -0.01 | -0.16 | 0.08 | -0.07 |
| 4 | $10.227 \times 10^9$ | 0.04 | 0.02 | -0.18 | 0.12 | 0.27 | -0.23 | -0.28 | -0.15 | -0.25 | -0.56 | -0.29 |
| 5 | $7.692 \times 10^9$ | -0.13 | 0.05 | -0.46 | -0.17 | -0.21 | 0.28 | 0.12 | 0.21 | 0.51 | 0.67 | 0.46 |

## 6 Conclusion

We have defined set functions that show the relative features of alternatives. There is room for further research into the interpretation of these set functions. In the uniform generating method, the simulation number $K$ increases exponentially with $n$ or $d$, especially in Choquet integral cases. In Choquet integral cases, it is hard to give an intuitive explanation of the difference between $E^W_{X,j}(A)$ and $M^*_{X,j}(A)$ when $\mid A \mid \geq 2$.

## References


1. Shapley, L.: A value for n-person games. Contribution Theory of Games, II. Annals of Mathematics Studies 28, 307–317 (1953)
2. Choquet, G.: Theory of capacities. Annales de l'Institut Fourier 5, 131–295 (1954)
3. Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill (1980)
4. Murofushi, T., Sugeno, M.: A theory of fuzzy measure: Representation, the Choquet integral and null sets. J. Math. Anal. Appl. 159, 532–549 (1991)
5. Murofushi, T., Soneda, S.: Techniques for reading fuzzy measures (III): interaction index. In: 9th Fuzzy System Symposium, Sapporo, Japan, pp. 693–696 (1993) (in Japanese)
6. Klein, D.A.: Decision analytic intelligent systems: automated explanation and knowledge acquisition. Lawrence Erlbaum Associates, Psychology Press (1994)
7. Fujimoto, K., Murofushi, T.: Some Characterizations of the Systems Represented by Choquet and Multi-Linear Functionals through the use of Möbius Inversion. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 5, 547–561 (1997)
8. Montmain, J., Mauris, G., Akharraz, A.: Elucidation and decisional risk in a multi criteria decision based on a choquet integral aggregation: A cybernetic framework. International Journal of Multi-Criteria Decision Analysis 13, 239–258 (2005)
9. Labreuche, C.: A general framework for explaining the results of a multi-attribute preference model. Artificial Intelligence 175, 1410–1448 (2011)
10. Takahagi, E.: Feature Analysis of Alternatives: Set Function Representations and Interpretations. Journal of Japan Society for Fuzzy Theory and Intelligent Informatics 26, 827–833 (2013) (in Japanese)

# 2-additive Choquet Optimal Solutions in Multiobjective Optimization Problems

Thibaut Lust[1] and Antoine Rolland[2]

[1] Sorbonne Universités, UPMC Univ Paris 06, LIP6-CNRS, Paris, France
thibaut.lust@lip6.fr
[2] Université de Lyon, ERIC, France
antoine.rolland@univ-lyon2.fr

**Abstract.** In this paper, we propose a sufficient condition for a solution to be optimal for a 2-additive Choquet integral in the context of multiobjective combinatorial optimization problems. A 2-additive Choquet optimal solution is a solution that optimizes at least one set of parameters of the 2-additive Choquet integral. We also present a method to generate 2-additive Choquet optimal solutions of multiobjective combinatorial optimization problems. The method is experimented on some Pareto fronts and the results are analyzed.

**Keywords:** Choquet integral, Multicriteria decision making, Fuzzy measure, Multiobjective combinatorial optimization, $k$-additivity.

## 1 Introduction

Multiobjective combinatorial optimization problems (MOCO) aim at finding the Pareto optimal solutions among a combinatorial set of feasible solutions. A Pareto optimal solution is a solution that is not Pareto dominated by any other solutions; the set of all these solutions is named the Pareto optimal set (or the Pareto front, in the objective space). However, the set of all the Pareto optimal solutions can be huge, especially in the case of several objectives [1]. Therefore it is worth to study the set of solutions that optimize a specific function, for example a weighted sum, as it generally reduces the size of the set of interesting Pareto optimal solutions. In this latter case, it is well-known that the set of potential optimal solutions is the convex envelop of the feasible solutions set. In order to attain solutions located in the non-convex part of the feasible solutions set, other aggregation operators could be used as function to be optimized. In this paper, we will focus on a specific aggregation operator: the Choquet integral.

The Choquet integral [2] is one of the most powerful tools in multicriteria decision making [3, 4]. A Choquet integral can be seen as an integral on a non-additive measure (or capacity or fuzzy measure), that is an aggregation operator that can model interactions between criteria. It presents extremely wide expressive capabilities and can model many specific aggregation operators, including, but not limited to, the weighted sum, the minimum, the maximum, all

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 256–265, 2014.

the statistic quantiles, the ordered weighted averaging operator [5], the weighted ordered weighted averaging operator [6], etc.

However, this high expressiveness capability needs a great number of parameters. While a weighted sum operator with $p$ criteria requires $p-1$ parameters, the definition of the Choquet integral with $p$ criteria requires the setting of $2^p-2$ values, which can be high even for low values of $p$. The notion of $k$-additivity has been introduced by Grabisch [7] in order to reduce the number of needed parameters while keeping the possibility to take into account the interactions between $k$ criteria among the $p$ criteria; typically for $k=2$, one only needs $\frac{p^2+p}{2}$ parameters.

Some papers already deal with the optimization of the Choquet integral of MOCO problems [8–10] when the Choquet integral is completely defined by the decision maker. Recently, Lust and Rolland investigated a method to generate the whole set of Choquet optimal solutions. The aim is to compute all the solutions that are potentially optimal for at least one parameter set of the Choquet integral. This method was studied in the particular case of biobjective combinatorial optimization problems [11], and for the general case in [12]. A characterization of the Choquet optimal solutions through a set of weighted-sum optimal solutions has been stated.

In this contribution, we focus on the specific case of the 2-additive Choquet integral. In the next section, we recall the definition of the Choquet integral. We propose then a sufficient condition for a solution to be Choquet optimal with a 2-additive capacity. We finally present some experimental results where we study the difference between the exact set of Choquet optimal with a 2-additive capacity and the set obtained with the sufficient condition proposed in this paper.

## 2 Aggregation Operators and Choquet Integral

We introduce in this section the basic concepts linked to multiobjective combinatorial optimization problems, the weighted sum and the Choquet integral.

### 2.1 Multiobjective Combinatorial Optimization Problems

A multiobjective (linear) combinatorial optimization (MOCO) problem is generally defined as follows:

$$\begin{aligned} &\text{“}\max_x\text{”} f(x) = Cx = (f_1(x), f_2(x), \ldots, f_p(x)) \\ &\text{subject to } Ax \leq b \\ &\qquad x \in \{0,1\}^n \end{aligned}$$

$$\begin{aligned} x \in \{0,1\}^n &\longrightarrow n \text{ variables} \\ C \in \mathbb{R}^{p\times n} &\longrightarrow p \text{ objective functions} \\ A \in \mathbb{R}^{r\times n} \text{ and } b \in \mathbb{R}^{r\times 1} &\longrightarrow r \text{ constraints} \end{aligned}$$

A feasible solution $x$ is a vector of $n$ variables, having to satisfy the $r$ constraints of the problem. Therefore, the feasible set in decision space is given by $\mathcal{X} = \{x \in \{0,1\}^n : Ax \leq b\}$. The image of the feasible set is given by $\mathcal{Y} = f(\mathcal{X}) = \{f(x) : x \in \mathcal{X}\} \subset \mathbb{R}^p$. An element of the set $\mathcal{Y}$ is called a cost-vector or a point.

Let us consider in the following, without loss of generality, that all the objectives have to be maximized and we design by $\mathcal{P}$ the set of objectives $\{1, \ldots, p\}$. The Pareto dominance relation ($P$-dominance for short) is defined, for all $y^1, y^2 \in \mathbb{R}^p$, by:

$$y^1 \succ_P y^2 \Longleftrightarrow [\forall k \in \mathcal{P}, y^1_k \geq y^2_k \text{ and } y^1 \neq y^2]$$

The strict Pareto dominance relation ($sP$-dominance for short) is defined as follows:

$$y^1 \succ_{sP} y^2 \Longleftrightarrow [\forall k \in \mathcal{P}, y^1_k > y^2_k]$$

Within a feasible set $\mathcal{X}$, any element $x^1$ is said to be *$P$-dominated* when $f(x^2) \succ_P f(x^1)$ for some $x^2$ in $\mathcal{X}$, *$P$-optimal* (or *$P$-efficient*) if there is no $x^2 \in \mathcal{X}$ such that $f(x^2) \succ_P f(x^1)$ and *weakly $P$-optimal* if there is no $x^2 \in \mathcal{X}$ such that $f(x^2) \succ_{sP} f(x^1)$. The *$P$-optimal set* denoted by $\mathcal{X}_P$ contains all the $P$-optimal solutions. The image $f(x)$ in the objective space of a $P$-optimal solution $x$ is called a $P$-non-dominated point. The image of the $P$-optimal set in $\mathcal{Y}$, equal to $f(\mathcal{X}_P)$, is called the Pareto front, and is denoted by $\mathcal{Y}_P$.

### 2.2 Weighted Sum

Instead of generating the *$P$-optimal set*, one can generate the solutions that optimize an aggregation operator. One of the most popular aggregation operator is the weighted sum (WS), where non-negative importance weights $\lambda_i (i = 1, \ldots, p)$ are allocated to the objectives.

**Definition 1.** *Given a vector $y \in \mathbb{R}^p$ and a weight set $\lambda \in \mathbb{R}^p$ (with $\lambda_i \geq 0$ and $\sum_{i=1}^p \lambda_i = 1$), the WS $f^{ws}_\lambda(y)$ of $y$ is equal to:*

$$f^{ws}_\lambda(y) = \sum_{i=1}^{p} \lambda_i y_i$$

**Definition 2.** *Let $x \in \mathcal{X}$ and $y = f(x)$ be its image in $\mathcal{Y}$. If $\exists\, \lambda \in \mathbb{R}^p_+$ $(\lambda_i > 0)$ such that $f^{ws}_\lambda(y) \geq f^{ws}_\lambda(y^2)\ \forall\, y^2 \in \mathcal{Y}$ then $x$ is a* supported *$P$-optimal solution, and its image $y$ a supported $P$-non-dominated point.*

Note that there exist $P$-optimal solutions that do not optimize a WS, and they are generally called *non-supported* $P$-optimal solutions [1].

### 2.3 Choquet Integral

The Choquet integral has been introduced by Choquet [2] in 1953 and has been intensively studied, especially in the field of multicriteria decision analysis, by several authors (see [3, 4, 13] for a brief review).

We first define the notion of capacity, on which the Choquet integral is based.

**Definition 3.** *A capacity is a set function* $v\colon 2^{\mathcal{P}} \to [0,1]$ *such that:*

- $v(\emptyset) = 0$, $v(\mathcal{P}) = 1$ *(boundary conditions)*
- $\forall \mathcal{A}, \mathcal{B} \in 2^{\mathcal{P}}$ *such that* $\mathcal{A} \subseteq \mathcal{B}, v(\mathcal{A}) \leq v(\mathcal{B})$ *(monotonicity conditions)*

Therefore, for each subset of objectives $\mathcal{A} \subseteq \mathcal{P}$, $v(\mathcal{A})$ represents the importance of the coalition $\mathcal{A}$.

**Definition 4.** *The Choquet integral of a vector* $y \in \mathbb{R}^p$ *with respect to a capacity* $v$ *is defined by:*

$$\begin{aligned} f_v^C(y) &= \sum_{i=1}^{p} \big(v(Y_i^{\uparrow}) - v(Y_{i+1}^{\uparrow})\big) y_i^{\uparrow} \\ &= \sum_{i=1}^{p} (y_i^{\uparrow} - y_{i-1}^{\uparrow}) v(Y_i^{\uparrow}) \end{aligned}$$

*where* $y^{\uparrow} = (y_1^{\uparrow}, \ldots, y_p^{\uparrow})$ *is a permutation of the components of* $y$ *such that* $0 = y_0^{\uparrow} \leq y_1^{\uparrow} \leq \ldots \leq y_p^{\uparrow}$ *and* $Y_i^{\uparrow} = \{j \in \mathcal{P}, y_j \geq y_i^{\uparrow}\} = \{i^{\uparrow}, (i+1)^{\uparrow}, \ldots, p^{\uparrow}\}$ *for* $i \leq p$ *and* $Y_{(p+1)}^{\uparrow} = \emptyset$.

We can notice that the Choquet integral is an increasing function of its arguments.

We can also define the Choquet integral through the Möbius representation [14] of the capacity. Any set function $v\colon 2^{\mathcal{P}} \to [0,1]$ can be uniquely expressed in terms of its Möbius representation by:

$$v(\mathcal{A}) = \sum_{\mathcal{B} \subseteq \mathcal{A}} m_v(\mathcal{B}) \quad \forall \mathcal{A} \subseteq \mathcal{P}$$

where the set function $m_v : 2^{\mathcal{P}} \to \mathbb{R}$ is called the Möbius transform or Möbius representation of $v$ and is given by

$$m_v(\mathcal{A}) = \sum_{\mathcal{B} \subseteq \mathcal{A}} (-1)^{(a-b)} v(\mathcal{B}) \quad \forall \mathcal{A} \subseteq \mathcal{P}$$

where $a$ and $b$ are the cardinals of $\mathcal{A}$ and $\mathcal{B}$.

A set of $2^p$ coefficients $m_v(\mathcal{A})$ $(\mathcal{A} \subseteq \mathcal{P})$ corresponds to a capacity if it satisfies the boundary and monotonicity conditions [15]:

1. $m_v(\emptyset) = 0$, $\sum_{\mathcal{A} \subseteq \mathcal{P}} m_v(\mathcal{A}) = 1$
2. $\sum_{\mathcal{B} \subseteq \mathcal{A},\ i \in \mathcal{B}} m_v(\mathcal{B}) \geq 0 \quad \forall \mathcal{A} \subseteq \mathcal{P}, i \in \mathcal{P}$

We can now write the Choquet integral with the use of Möbius coefficients. The Choquet integral of a vector $y \in \mathbb{R}^p$ with respect to a capacity $v$ is defined as follows:

$$f_v^C(y) = \sum_{\mathcal{A} \subseteq \mathcal{P}} m_v(\mathcal{A}) \min_{i \in \mathcal{A}} y_i$$

A Choquet integral is a versatile aggregation operator, as it can express preferences to a wider set of solutions than a weighted sum, through the use of a non-additive capacity. When solving a MOCO problem, with the Choquet integral, one can attain *non-supported* $P$-optimal solutions, while it is impossible with the weighted sum [11].

However, this model needs also a wider set of parameters to capture this non-additivity. For $p$ criteria, one only needs $p-1$ weights to use a weighted sum, where $2^p-2$ weights are needed to use a Choquet integral based on a capacity. Therefore, the concept of $k$-additivity has been introduced by [7] to find a compromise between the expressiveness of the model and the number of needed parameters.

**Definition 5.** *A capacity $v$ is said to be $k$-additive if*

- $\forall \mathcal{A} \subseteq \mathcal{P}, m_v(\mathcal{A}) = 0$ *if* $card(\mathcal{A}) > k$
- $\exists\, \mathcal{A} \subseteq \mathcal{P}$ *such that* $card(\mathcal{A}) = k$ *and* $m_v(\mathcal{A}) \neq 0$

We will specially focus in this paper on 2-additive capacities and propose a sufficient condition for a solution of a MOCO problem to be 2-additive Choquet optimal.

## 3 Characterization of Choquet Optimal Solutions

### 3.1 Choquet Optimal Solutions

A characterization of the set of Choquet optimal solutions of a MOCO problem has been proposed in [12]. We briefly recall it here.

We denote the set of Choquet optimal solutions of a MOCO problem with $p$ objectives $\mathcal{X}_C$: it contains at least one solution $x \in \mathcal{X}$ optimal for each possible Choquet integral, that is $\forall v \in \mathcal{V}, \exists\, x_c \in \mathcal{X}_C \,|\, f_v^C(f(x_c)) \geq f_v^C(f(x))\ \forall x \in \mathcal{X}$, where $\mathcal{V}$ represents the set of capacity functions defined over $p$ objectives. Note that each Choquet optimal solution is at least weakly $P$-optimal [11].

In [11], Lust and Rolland studied the particular case of two objectives and they showed that $\mathcal{X}_C$ could be obtained by generating all WS-optimal solutions in each subspace of the objectives separated by the bisector ($f_1(x) \geq f_2(x)$ or $f_2(x) \geq f_1(x)$), and by adding a particular point $M$ with $M_1 = M_2 = \max_{x \in \mathcal{X}} \min(f_1(x), f_2(x))$.

In [12], Lust and Rolland extended this characterization to the general case.

Let $\sigma$ be a permutation on $\mathcal{P}$. Let $O_\sigma$ be the subset of points $y \in \mathbb{R}^p$ such that $y \in O_\sigma \iff y_{\sigma_1} \geq y_{\sigma_2} \geq \ldots \geq y_{\sigma_p}$.

Let $p_{O_\sigma}$ be the following application:

$$p_{O_\sigma} : \mathbb{R}^p \to \mathbb{R}^p, (p_{O_\sigma}(y))_{\sigma_i} = (\min(y_{\sigma_1}, \ldots, y_{\sigma_i})), \forall i \in \mathcal{P}$$

For example, if $p = 3$, for the permutation (2,3,1), we have:

$$p_{O_\sigma}(y) = \big(\min(y_2, y_3, y_1), \min(y_2), \min(y_2, y_3)\big)$$

We denote by $\mathcal{P}_{O_\sigma}(\mathcal{Y})$ the set containing the points obtained by applying the application $p_{O_\sigma}(y)$ to all the points $y \in \mathcal{Y}$. As $(p_{O_\sigma}(y))_{\sigma_1} \geq (p_{O_\sigma}(y))_{\sigma_2} \geq \ldots \geq (p_{O_\sigma}(y))_{\sigma_p}$, we have $\mathcal{P}_{O_\sigma(\mathcal{Y})} \subseteq O_\sigma$.

**Theorem 1**

$$\mathcal{Y}_C \cap O_\sigma = \mathcal{Y} \cap WS(\mathcal{P}_{O_\sigma}(\mathcal{Y}))$$

where $WS(\mathcal{P}_{O_\sigma}(\mathcal{Y}))$ designs the set of WS-optimal points of the set $\mathcal{P}_{O_\sigma}(\mathcal{Y})$.

This theorem characterizes the solutions which can be Choquet optimal in the set of feasible solutions as being, in each subspace of the objective space $\mathcal{Y}$ where $y_{\sigma_1} \geq y_{\sigma_2} \geq \ldots \geq y_{\sigma_p}$, the solutions that have an image corresponding to a WS-optimal point in the space composed of the original subspace plus the projection of all the other points following the application $p_{O_\sigma}$.

**Proof:** see [12].

### 3.2 2-additive Choquet Optimal Solutions

We are now interested in the definition of the set of solutions of a MOCO problem that potentially optimize a 2-additive Choquet integral (and not a general Choquet integral). How does the constraints of 2-additivity restrict the set $\mathcal{Y}_C$? We will denote $\mathcal{Y}_{C^2}$ the set of 2-additive Choquet optimal solutions. As stated above, $\sigma$ is a permutation on $\mathcal{P}$ and $O_\sigma$ is the subset of points $y \in \mathbb{R}^p$ such that $y \in O_\sigma \iff y_{\sigma_1} \geq y_{\sigma_2} \geq \ldots \geq y_{\sigma_p}$.

**Theorem 2**

$$\forall \delta \in \Delta, \textit{if } y \in \mathcal{Y} \cap WS(\mathcal{P}^\delta_\sigma(\mathcal{Y})) \Rightarrow y \in \mathcal{Y}_{C^2} \cap O_\sigma$$

where:

- $\delta$ is an application $\mathcal{P} \to \mathcal{P}$ such that $\delta(1) = 1$ and $\delta(i) < i \ \forall i \neq 1$. Let $\Delta$ be the set of all applications $\delta$.
- $p^\delta_\sigma$ is an application on $\mathcal{Y}$ such that $(p^\delta_\sigma(y))_{\sigma_i} = \min(y_{\sigma_{\delta(i)}}, y_{\sigma_i})$.
  For example, if $p = 4$, for the permutation (1,2,3,4) and $\delta = (1, 1, 2, 3)$, we have:
  - $(p^\delta_\sigma(y))_1 = \min(y_1, y_1) = y_1$
  - $(p^\delta_\sigma(y))_2 = \min(y_1, y_2)$
  - $(p^\delta_\sigma(y))_3 = \min(y_2, y_3)$
  - $(p^\delta_\sigma(y))_4 = \min(y_3, y_4)$
- $\mathcal{P}^\delta_\sigma(\mathcal{Y})$ is the set containing the points obtained by applying the application $p^\delta_\sigma$ to all the points $y \in \mathcal{Y}$.
- $WS(\mathcal{P}^\delta_\sigma(\mathcal{Y}))$ designs the set of supported points of the set $\mathcal{P}^\delta_\sigma(\mathcal{Y})$.

**Proof**

In the following, we will denote $O_\sigma$ as simply $O$ for the sake of simplicity, and we will consider, without loss of generality, that the permutation $\sigma$ is equal to $(1, 2, \ldots, p)$, that is $y \in O \Leftrightarrow y_1 \geq y_2 \geq \cdots \geq y_p$. We will consequently note $p^\delta_\sigma$ as simply $p^\delta$ and $\mathcal{P}^\delta_\sigma(\mathcal{Y})$ as $\mathcal{P}^\delta(\mathcal{Y})$. We know that $\mathcal{Y}_{C^2} \subseteq \mathcal{Y}$ and then $\mathcal{Y}_{C^2} \cap O \subseteq \mathcal{Y} \cap O$.

Let us suppose that $y \in O$. Let $y \in WS(\mathcal{P}^\delta(\mathcal{Y})) \cap \mathcal{Y}$. Then there are $\lambda_1, \ldots, \lambda_p \geq 0$ such that $\sum_{i=1}^{p} \lambda_i = 1$ and

$$\forall z \in \mathcal{Y},\ \sum_{i \in \mathcal{P}} \lambda_i y_i \geq \sum_{i \in \mathcal{P}} \lambda_i p^\delta(z)_i$$

By definition, $p^\delta(z)_i = \min(z_{\delta(i)}, z_i), \forall i \in \mathcal{P}$.

Let $\mathcal{A} \subseteq \mathcal{P}$. Let us define a set function $m$ such that $m(\mathcal{A}) = \lambda_i$ if $\mathcal{A} = \{\delta(i), i\}$ and $m(\mathcal{A}) = 0$ if not.

Then

$$\begin{aligned} \sum_{i \in \mathcal{P}} \lambda_i (p^\delta(z))_i &= \sum_{i \in \mathcal{P}} \lambda_i \min(z_{\delta(i)}, z_i) \\ &= \sum_{\mathcal{A} \subseteq \mathcal{P}} m(\mathcal{A}) \min_{i \in \mathcal{A}} z_i \end{aligned}$$

Let us remind that the set function $m$ corresponds to a capacity $v$ if:

1. $m(\emptyset) = 0,\ \sum_{\mathcal{A} \subseteq \mathcal{P}} m(\mathcal{A}) = 1$
2. $\sum_{\mathcal{B} \subseteq \mathcal{A},\ i \in \mathcal{B}} m(\mathcal{B}) \geq 0 \quad \forall \mathcal{A} \subseteq \mathcal{P}, i \in \mathcal{P}$

All these conditions are satisfied:

- $m(\emptyset) = 0$ by definition
- $\sum_{\mathcal{A} \subseteq \mathcal{P}} m(\mathcal{A}) = \sum_{i=1}^{p} \lambda_i = 1$
- all $m(\mathcal{B})$ are non-negative as $\lambda_i \geq 0$

Moreover, as $m(\mathcal{A}) = 0\ \forall \mathcal{A}$ such that $card(\mathcal{A}) > 2$, $v$ is a 2-additive capacity. Therefore we have a capacity $v$ and its set of Möbius coefficients such that

$\forall z \in \mathcal{Y}$,

$$\begin{aligned} f_v^C(y) &= \sum_{\mathcal{A} \subseteq \mathcal{P}} m(\mathcal{A}) \min_{i \in \mathcal{A}} y_i \\ &= \sum_{i \in \mathcal{P}} \lambda_i y_i \\ &\geq \sum_{i \in \mathcal{P}} \lambda_i p^\delta(z)_i \\ &\geq \sum_{\mathcal{A} \subseteq \mathcal{P}} m(\mathcal{A}) \min_{i \in \mathcal{A}} z_i \\ &\geq f_v^C(z) \end{aligned}$$

So $y \in \mathcal{Y}_{C^2}$. □

From this theorem, we can derive an algorithm to generate a set $\mathcal{X}_{C^2}$ containing solutions of a MOCO problem that optimize a 2-additive Choquet integral.

For all the permutations $\sigma$ on $\mathcal{P}$, we have to:

1. Consider an application $\delta$ such that $\delta(1) = 1$ and $\delta(i) < i \ \forall i \neq 1$.
2. Determine the set $\mathcal{P}_\sigma^\delta(\mathcal{Y})$ containing the projections obtained with the application $p_\sigma^\delta$ for each $y \in \mathcal{Y}$.
3. Determine the solutions in $O_\sigma$ that optimize a WS considering $\mathcal{P}_\sigma^\delta(\mathcal{Y})$.

## 4 Experiments

We have applied the algorithm for defined Pareto fronts, that is, a Pareto front is given, and the aim is to determine, among the $P$-non-dominated points, the 2-additive Choquet optimal points.

To generate Pareto fronts, we have applied a heuristic to multiobjective knapsack instances. We have used knapsack instances with random profits. The heuristic is an adaptation of the one presented in [16]. Note that the aim is only to generate a set of non-dominated points to experiment the sufficient condition.

The results are given in Table 1 for $p = 4$, $k = 2$, and 250 points, and in Table 2 for $p = 4$, $k = 2$, and 500 points.

We have considered all possible applications $\delta$. We have also computed the exact number of 2-additive Choquet optimal solutions with a linear program: for each point of the Pareto front, we check if there exists a 2-additive capacity $v$ such that the Choquet integral of this point is better that all the other points. Note that this method can be applied since we consider the particular case of a given Pareto front. For the problem with 250 points, 140 points optimize a 2-additive Choquet integral and for the problem with 500 points, 200 points optimize a 2-additive Choquet integral. We see that our method can only reach a subset of this set (since the method is only based on a sufficient condition). The number of 2-additive Choquet optimal points generated depends on the

**Table 1.** Random multiobjective knapsack instances (250 points, 140 are 2-additive Choquet optimal)

| $\delta$ | #2$C$-Optimal |
|---|---|
| (1,1,1,1) | 124 |
| (1,1,1,2) | 124 |
| (1,1,1,3) | 131 |
| (1,1,2,1) | 131 |
| (1,1,2,2) | 131 |
| (1,1,2,3) | 139 |

**Table 2.** Random multiobjective knapsack instances (500 points, 200 are 2-additive Choquet optimal)

| $\delta$ | #2$C$-Optimal |
|---|---|
| (1,1,1,1) | 164 |
| (1,1,1,2) | 164 |
| (1,1,1,3) | 174 |
| (1,1,2,1) | 186 |
| (1,1,2,2) | 187 |
| (1,1,2,3) | 196 |

application $\delta$. For the set with 250 points, with $\delta = (1, 1, 1, 1)$, 124 points are generated, while with the application $(1, 1, 2, 3)$, 139 points are computed. For the set with 500 points, with $\delta = (1, 1, 1, 1)$, 164 points are generated, while with the application $(1, 1, 2, 3)$, 196 points are computed. Some application $\delta$ allows thus to reach more 2-additive Choquet optimal solutions. However, even by merging the sets obtained with all possible applications $\delta$, they are still 2-additive Choquet optimal solutions that cannot be reached with our method based on the sufficient condition.

## 5 Conclusion

We have introduced in this paper a sufficient condition to produce 2-additive Choquet optimal solutions of multiobjective combinatorial optimization problems. We have also presented an algorithm to obtain these solutions based on this condition. The algorithm can be applied to generate an interesting subset of the Pareto optimal set (in case of the size of this set is too high). This work about generating 2-additive Choquet optimal solutions opens many new perspectives:

– As our condition is only sufficient, a necessary and sufficient condition will be required to generate all the 2-additive Choquet optimal solutions of a MOCO problem. The condition will also have to be generalized to $k > 2$.

- Following [17], it will be interesting to study and to define what brings exactly and concretely (for a decision maker) the 2-additive Choquet optimal solutions that are not WS optimal solutions, given that they are harder to compute.
- More experiments will be needed to show the differences between WS optimal solutions, Choquet optimal solutions and 2-additive Choquet optimal solutions of MOCO problems.

## References


1. Ehrgott, M.: Multicriteria Optimization, 2nd edn. Springer, Berlin (2005)
2. Choquet, G.: Theory of capacities. Annales de l'Institut Fourier (5), 131–295 (1953)
3. Grabisch, M., Labreuche, C.: A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. Annals OR 175(1), 247–286 (2010)
4. Grabisch, M., Marichal, J.C., Mesiar, R., Pap, E.: Aggregation functions. Encyclopedia of Mathematics and its Applications, vol. 127. Cambridge University Press, Cambridge (2009)
5. Yager, R.: On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Trans. Systems, Man and Cybern. 18, 183–190 (1998)
6. Torra, V.: The weighted OWA operator. Int. J. Intell. Syst. (12), 153–166 (1997)
7. Grabisch, M.: k-order additive discrete fuzzy measures and their representation. Fuzzy Sets and Systems (92), 167–189 (1997)
8. Fouchal, H., Gandibleux, X., Lehuédé, F.: Preferred solutions computed with a label setting algorithm based on Choquet integral for multi-objective shortest paths. In: 2011 IEEE Symposium on Computational Intelligence in Multicriteria Decision-Making (MDCM), pp. 143–150. IEEE, Paris (2011)
9. Galand, L., Perny, P., Spanjaard, O.: Choquet-based optimisation in multiobjective shortest path and spanning tree problems. European Journal of Operational Research 204(2), 303–315 (2010)
10. Galand, L., Perny, P., Spanjaard, O.: A branch and bound algorithm for Choquet optimization in multicriteria problems. Proceedings Lecture Notes in Economics and Mathematical Systems 634, 355–365 (2011)
11. Lust, T., Rolland, A.: Choquet optimal set in biobjective combinatorial optimization. Computers & Operations Research 40(10), 2260–2269 (2013)
12. Lust, T., Rolland, A.: On the computation of Choquet optimal solutions in multicriteria decision contexts. In: Ramanna, S., Lingras, P., Sombattheera, C., Krishna, A. (eds.) MIWAI 2013. LNCS, vol. 8271, pp. 131–142. Springer, Heidelberg (2013)
13. Grabisch, M.: The application of fuzzy integrals in multicriteria decision making. European Journal of Operational Research 89, 445–456 (1996)
14. Rota, G.C.: On the foundations of combinatorial theory I. Theory of Mobius functions. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 2, 340–368 (1964)
15. Chateauneuf, A., Jaffray, J.Y.: Some characterizations of lower probabilities and other monotone capacities through the use of mobius inversion. Mathematical Social Sciences 17(3), 263–283 (1989)
16. Lust, T., Teghem, J.: The multiobjective multidimensional knapsack problem: a survey and a new approach. International Transactions in Operational Research 19(4), 495–520 (2012)
17. Meyer, P., Pirlot, M.: On the expressiveness of the additive value function and the Choquet integral models. In: DA2PL 2012: from Multiple Criteria Decision Aid to Preference Learning, pp. 48–56 (2012)

# A Characterization of the 2-Additive Symmetric Choquet Integral Using Trinary Alternatives

Brice Mayag

LAMSADE, University Paris-Dauphine
Place du Marchal de Lattre de Tassigny
75775 Paris cedex 16 France
brice.mayag@dauphine.fr
http://www.lamsade.dauphine.fr/~mayag/

**Abstract.** In a context of Multiple Criteria Decision Aid, we present some necessary and sufficient conditions to obtain a symmetric Choquet integral compatible with some preferences on a particular set of alternatives. These axioms are based on the notion of strict cycle and the MOPI conditions.

**Keywords:** Choquet integral, binary alternatives, capacity, trinary alternatives.

## 1 Introduction

Multiple criteria decision Aid (MCDA) aims at representing the preferences of a decision maker (DM) over a set of options (or alternatives) $X = X_1 \times \cdots \times X_n$, in regard to a finite set of criteria $N = \{1, \ldots, n\}$ or attributes $X_1, \ldots, X_n$. One possible model is the transitive decomposable one where an overall utility is determined for each alternative. The 2-additive Choquet integral has been proved to be a versatile aggregation function to construct overall scores [3,4,12,13] and is based on the notion of the 2-additive capacity or fuzzy measure. This model assumes that partial utilities belong to non-negative or unipolar scales.

Grabisch and Labreuche [5,6] and Bouyssou et al [2] show through some motivating examples that, sometimes, unipolar scales are not appropriate to represent the DM's preferences. Therefore, in some situations, bipolar scales appear more useful. Bipolar scale is defined as a scale composed of a negative, a positive and a neutral part which respectively allow representing a negative, a positive and a neutral affect towards an option. The Choquet integral has been extended to the bipolar scale based on the notion of capacity or on the notion of bi-capacity [5,8,9]. One of these extensions is the symmetric Choquet integral, also called Šipoš integral, defined by a capacity. In this paper we consider a symmetric Choquet integral defined by a 2-additive capacity.

To identify a 2-additive capacity for symmetric Choquet integral, we assume that the DM can provided an ordinal information on a particular set of alternatives, the set of trinary actions. An ordinal information is a preference information represented by a strict preference relation and an indifference relation.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 266–275, 2014.

A trinary action is a fictitious alternative which takes either the neutral value **0** for all criteria, or the neutral value **0** for all criteria except for one or two criteria for which it takes the satisfactory value **1** or the unsatisfactory value $-\mathbf{1}$. We think that this hypothesis is reasonable due to the bipolar scale that we manipulate. We study the case when this ordinal information on trinary actions can be considered as an ordinal information on unipolar scales.

We present the necessary and sufficient conditions on this ordinal information for the existence of a 2-additive capacity such that the symmetric Choquet integral w.r.t. this capacity represents the preference of the decision maker. The first condition concerns the existence of strict cycles in the ordinal information, the second condition concerns the strict cyles in the ordinal information restricted to unipolar scales and the last condition, called MOPI condition, comes from the definition of a 2-additive capacity.

The next section show how to represent an ordinal information by a 2-additive Choquet integral, while the last section how to represent an ordinal information in the case of symmetric Choquet integral.

## 2 The Representation of the Ordinal Information by a 2-additive Choquet Integral

The 2-additive Choquet integral is a particular case of the well known Choquet integral [10]. Its main property is to model interactions between two criteria. These interactions are simple and more meaningful than those produced by using the Choquet integral. This aggregation function is based on the notion of *capacity* $\mu$ defined as a set function from the powerset of criteria $2^N$ to $[0,1]$ such that:

1. $\mu(\emptyset) = 0$
2. $\mu(N) = 1$
3. $\forall A, B \in 2^N,\ [A \subseteq B \Rightarrow \mu(A) \leq \mu(B)]$ (monotonicity).

A capacity $\mu$ on $N$ is said to be *2-additive* if its *Möbius transform* $m : 2^N \to \mathbb{R}$ defined by

$$m(T) := \sum_{K \subseteq T} (-1)^{|T \setminus K|} \mu(K), \forall T \in 2^N. \tag{1}$$

satisfies the following two conditions:

- For all subset $T$ of $N$ such that $|T| > 2$, $m(T) = 0$;
- There exists a subset $B$ of $N$ such that $|B| = 2$ and $m(B) \neq 0$.

Given an alternative
$x := (x_1, ..., x_n) \in X$, the expression of the 2-additive Choquet integral is given by [6]:

$$C_\mu(u(x)) = \sum_{i=1}^{n} v_i u_i(x_i) - \frac{1}{2} \sum_{\{i,j\} \subseteq N} I_{ij} |u_i(x_i) - u_j(x_j)| \tag{2}$$

where

- For all $i \in N$, $u_i : X_i \to \mathbb{R}_+$ is an utility function associated to the attribute $X_i$;
- $u(x) = (u_1(x_1), \ldots, u_n(x_n))$ for $x = (x_1, ..., x_n) \in X$;
- $v_i = \sum_{K \subseteq N \setminus i} \frac{(n - |K| - 1)!|K|!}{n!} (\mu(K \cup i) - \mu(K))$ is the importance of criterion $i$ corresponding to the Shapley value of $\mu$ [16];
- $I_{ij} = \mu_{ij} - \mu_i - \mu_j$ is the interaction index between the two criteria $i$ and $j$ [3,15].

Therefore the 2-additive Choquet integral appears as a good compromise between the arithmetic mean and the Choquet integral.

We simplify our notation for a capacity $\mu$ by using the following shorthand: $\mu_i := \mu(\{i\})$, $\mu_{ij} := \mu(\{i, j\})$ for all $i, j \in N$, $i \neq j$. Whenever we use $i$ and $j$ together, it always means that they are different.

### 2.1 The Ordinal Information on Binary Alternatives

We assume that the DM is able to identify for each criterion $i$ two reference levels:

1. A reference level $\mathbf{1}_i$ in $X_i$ which he considers as good and completely satisfying if he could obtain it on criterion $i$, even if more attractive elements could exist. This special element corresponds to the *satisficing level* in the theory of bounded rationality of [17].
2. A reference level $\mathbf{0}_i$ in $X_i$ which he considers neutral on $i$. The neutral level is an element which is thought by the DM to be neither good nor bad, neither attractive nor repulsive relatively to his concerns with respect to the criterion $i$. The existence of this neutral level has roots in psychology (see [18]), and is used in bipolar models like Cumulative Prospect Theory (see [19]).

We set for convenience $u_i(\mathbf{1}_i) = 1$ and $u_i(\mathbf{0}_i) = 0$. The use of Choquet integral requires to ensure the commensurateness between criteria. Therefore the previous reference levels can be used in order to define the same scale on each criterion [7,11]. For more details about these reference levels, see [4,6].

We call a *binary action or binary alternative*, an element of the set

$$\mathcal{B} = \{\mathbf{0}_N,\ (\mathbf{1}_i, \mathbf{0}_{N-i}),\ (\mathbf{1}_{ij}, \mathbf{0}_{N-ij}),\ i, j \in N, i \neq j\} \subseteq X$$

where

- $\mathbf{0}_N = (\mathbf{1}_\emptyset, \mathbf{0}_N) =: a_0$ is an action considered neutral on all criteria.
- $(\mathbf{1}_i, \mathbf{0}_{N-i}) =: a_i$ is an action considered satisfactory on criterion $i$ and neutral on the other criteria.
- $(\mathbf{1}_{ij}, \mathbf{0}_{N-ij}) =: a_{ij}$ is an action considered satisfactory on criteria $i$ and $j$ and neutral on the other criteria.

The number of binary actions is $1 + n + \frac{n \times (n-1)}{2} = 1 + \frac{n \times (n+1)}{2}$. For any 2-additive capacity $\mu$, we get the following consequences:

$$C_\mu(u(a_0)) = 0 \tag{3}$$
$$C_\mu(u(a_i)) = \mu_i \tag{4}$$
$$C_\mu(u(a_{ij})) = \mu_{ij} \tag{5}$$

Therefore, in order to compute all the parameters of the 2-additive Choquet integral, Mayag et al. [13] suggest to ask to the DM some preference information $\{P,\ I\}$ on the set of binary actions, called *ordinal information on* $\mathcal{B}$, and given by:

$$P = \{(x, y) \in \mathcal{B} \times \mathcal{B} :\ \text{DM strictly prefers } x \text{ to } y\} \tag{6}$$
$$I = \{(x, y) \in \mathcal{B} \times \mathcal{B} :\ \text{DM is indifferent between } x \text{ and } y\} \tag{7}$$

The following notations are used indifferently: $(x, y) \in P$ is equivalent to $x\ P\ y$ and $(x, y) \in I$ is equivalent to $x\ I\ y$.

To test if these preferences are representable by a 2-additive Choquet integral, a relation $M$ modelling the simple monotonicity conditions $\mu(\{i\}) \geq 0$ and $\mu(\{i, j\}) \geq \mu(\{i\})$ for a capacity $\mu$ is added to $\{P,\ I\}$. The relation $M$ is defined such that for $(x, y) \in \{(a_i, a_0), i \in N\} \cup \{(a_{ij}, a_i), i, j \in N, i \neq j\}$,

$$x\ M\ y \text{ if not}(x\ (P \cup I)\ y).$$

### 2.2 MOPI Conditions and the Characterization Theorem

Given an ordinal information $\{P, I\}$ on $\mathcal{B}$, Mayag et al. [13] give necessary and sufficient conditions on $\mathcal{B}$ for which $\{P, I\}$ is representable by a 2-additive Choquet integral i.e. for which there exists a 2-additive capacity $\mu$ such that:

$$(x, y) \in P \Longrightarrow C_\mu(u(x)) > C_\mu(u(y)) \tag{8}$$
$$(x, y) \in I \Longrightarrow C_\mu(u(x)) = C_\mu(u(y)) \tag{9}$$

This characterization is based on the MOPI property defined by:

**Definition 1.** *[MOPI property] Let $i, j, k \in N$, $i$ fixed.*

1. *A* Monotonicity of Preferential Information in $\{i, j, k\}$ w.r.t. $i$ *is the following property (denoted by* ***($\{i, j, k\}$,$i$)-MOPI****):*

$$\left.\begin{array}{l} a_{ij} \sim a_{i \vee j} \\ a_{ik} \sim a_{i \vee k} \\ i \vee j \neq i \vee k \end{array}\right\} \Rightarrow [not(a_l\ TC_P\ a_0), l \in \{i, j, k\} \setminus \{i \vee k, i \vee j\}] \tag{10}$$

*where*

- *The criterion $i \vee j$ refers to criterion $i$ or $j$.*
- *The relation $TC_P$ is defined by $x\ TC_P\ y$ if there exists a path of $(P \cup I \cup M)$ from $x$ to $y$ (i.e. there exist $x_1, \ldots, x_p$ such that $x = x_1\ (P \cup I \cup M)\ x_2\ (P \cup I \cup M) \cdots (P \cup I \cup M)\ x_{p-1}\ (P \cup I \cup M)\ x_p = y$) which is strict (i.e. there exists $i$ in $\{1, ..., p-1\}$ such that $x_i\ P\ x_{i+1}$).*
- *The relation $\sim$ on $\mathcal{B}$ is defined by $x \sim y$ if there exists a cycle of $(I \cup M)$ containing $x$ and $y$ (a path from $x$ to $y$ and a path from $y$ and $x$).*

2. *The set $\{i, j, k\}$ satisfies the property of* MOnotonicity of Preferential Information *(MOPI) if $\forall l \in \{i, j, k\}$, ($\{i, j, k\}$,$l$)-MOPI is satisfied.*

**Example 1.** *Let $N = \{1, 2, 3, 4\}$ and $i = 1$ fixed. The property $(\{1, 2, 3\}, 1)$-MOPI reads as follows:*

$$\begin{cases} a_{12} \sim a_2 \\ a_{13} \sim a_1 \end{cases} \Rightarrow not(a_3\ TC_P\ a_0)$$
*and*
$$\begin{cases} a_{12} \sim a_1 \\ a_{13} \sim a_3 \end{cases} \Rightarrow not(a_2\ TC_P\ a_0)$$
*and*
$$\begin{cases} a_{12} \sim a_2 \\ a_{13} \sim a_3 \end{cases} \Rightarrow not(a_1\ TC_P\ a_0).$$

Now we suppose $P \neq \emptyset$ ("non triviality" axiom).

**Theorem 1.** *An ordinal information $\{P, I\}$ is representable by a 2-additive Choquet integral on $\mathcal{B}$ if and only if the following conditions are satisfied:*

1. *$(P \cup I \cup M)$ contains no strict cycle;*
2. *Any subset $K$ of $N$ such that $|K| = 3$ satisfies the MOPI property.*

*Proof.* See [13].

## 3 The Represention of the Ordinal Information by a 2-additive Symmetric Choquet Integral

The symmetric Choquet integral, also called Šipoš integral, is a simple way to generalize Choquet integral for bipolar scales [5,8,9]. Given an alternative $x := (x_1, ..., x_n) \in X$ and a 2-additive capacity, the expression of the symmetric Choquet integral is given by:

$$\breve{C}_\mu(\breve{u}(x)) = C_\mu(\breve{u}(x)^+) - C_\mu(\breve{u}(x)^-) \tag{11}$$

where

- $\breve{u}(x) = (\breve{u}_1(x_1), \ldots, \breve{u}_n(x_n))$ with $\breve{u}_i : X_i \to \mathbb{R}$ the utility function associated to $X_i$, $i = 1, \ldots, n$;
- $\breve{u}(x)^+ := (\breve{u}_1(x_1)^+, \ldots, \breve{u}_n(x_n)^+)$ with $\breve{u}_i(x_i)^+ = \breve{u}_i(x_i) \vee 0$, $i = 1, \ldots, n$;
- $\breve{u}(x)^- := (\breve{u}_1(x_1)^-, \ldots, \breve{u}_n(x_n)^-)$ with $\breve{u}_i(x_i)^- = (-\breve{u}_i(x_i)) \vee 0$, $i = 1, \ldots, n$.

We call $\breve{C}_\mu$ a 2-additive symmetric Choquet integral when $\mu$ is 2-additive. Because this aggregation function uses bipolar scales, it can be also defined from a symmetric bi-capacity, see [9] for more details about this concept. We just recall that a function $\nu : 3^N \to \mathbb{R}$ is a bi-capacity on $3^N$ if it satisfies the following two conditions :

$$\nu(\emptyset, \emptyset) = 0 \tag{12}$$

$$\forall (A_1, A_2), (B_1, B_2) \in 3^N : [(A_1, A_2) \sqsubseteq (B_1, B_2) \Rightarrow \nu(A_1, A_2) \leq \nu(B_1, B_2)] \tag{13}$$

Where

- $3^N := \{(A, B) \in 2^N \times 2^N | A \cap B = \emptyset\}$ is the set of couples of subsets of $N$ with an empty intersection;
- $(A_1, A_2) \sqsubseteq (B_1, B_2) \Leftrightarrow [A_1 \subseteq B_1 \text{ and } B_2 \subseteq A_2]$.

### 3.1 The Ordinal Information on Trinary Alternatives

In addition to the two reference levels $\mathbf{1}_i$ and $\mathbf{0}_i$ in $X_i$, we assume that the DM is also able to identify for each criterion $i$ the reference level $-\mathbf{1}_i$ in $X_i$. This corresponds to the level which he considers completely unsatisfying. We set $\breve{u}_i(-\mathbf{1}_i) = -1$. Using these three reference levels, we consider now the set *trinary alternatives or trinary actions*[1] denoted by $\mathcal{T}$ and defined by

$$\mathcal{T} = \{(\mathbf{1}_\emptyset, -\mathbf{1}_\emptyset),\ (\mathbf{1}_i, -\mathbf{1}_\emptyset),\ (\mathbf{1}_\emptyset, -\mathbf{1}_j),\ (\mathbf{1}_i, -\mathbf{1}_j),\ (\mathbf{1}_{ij}, -\mathbf{1}_\emptyset),\ (\mathbf{1}_\emptyset, -\mathbf{1}_{ij}),\ i, j \in N\} \subseteq X,$$

where

- $(\mathbf{1}_\emptyset, -\mathbf{1}_\emptyset) =: a_{0|0}$ is an action considered neutral on all criteria.
- $(\mathbf{1}_i, -\mathbf{1}_\emptyset) =: a_{i|}$ is an action considered satisfactory on criterion $i$ and neutral on the other criteria.
- $(\mathbf{1}_\emptyset, -\mathbf{1}_j) =: a_{|j}$ is an action considered unsatisfactory on criterion $j$ and neutral on the other criteria.
- $(\mathbf{1}_i, -\mathbf{1}_j) =: a_{i|j}$ is an action considered satisfactory on criteria $i$, unsatisfactory on $j$ and neutral on the other criteria.
- $(\mathbf{1}_{ij}, -\mathbf{1}_\emptyset) := a_{ij|}$ is an action considered satisfactory on criteria $i$ and $j$ and neutral on the other criteria.
- $(\mathbf{1}_\emptyset, -\mathbf{1}_{ij}) := a_{|ij}$ is an action considered unsatisfactory on criteria $i$ and $j$ and neutral on the other criteria.

The trinary actions are used in [14] to elicitate a bi-capacity and are adapted for bipolar scales. Their number is $1 + 2 \times n + \dfrac{2 \times n \times (n-1)}{2} = 1 + 2 \times n^2$. Roughly speaking there are 4 times as much trinary actions for 2 additive bi-capacities compare to the 2 additive capacities.

[1] Some authors, like in [1], prefer the term "ternary alternatives" instead of trinary alternatives.

**Example 2.** *If* $N = \{1,2,3\}$ *then* $\mathcal{T} = \{a_{0|0}; a_{1|}; a_{2|}; a_{3|}; a_{|1}; a_{|2}; a_{|3}; a_{1|2}; a_{1|3}; a_{2|1};\ a_{2|3}; a_{3|1};\ a_{3|2}; a_{12|}; a_{13|}; a_{|12}; a_{23|}; a_{|13}; a_{|23}\}$

Given a 2-additive capacity $\mu$, we get the following consequences for a symmetric Choquet integral:

$$\breve{C}_\mu(\breve{u}(a_{0|0})) = C_\mu(u(a_0)) = 0 \tag{14}$$

$$\breve{C}_\mu(\breve{u}(a_{i|})) = C_\mu(\breve{u}(a^+_{i|})) - C_\mu(\breve{u}(a^-_{i|})) = \mu_i = C_\mu(u(a_i)) \tag{15}$$

$$\breve{C}_\mu(\breve{u}(a_{|j})) = C_\mu(\breve{u}(a^+_{|j})) - C_\mu(\breve{u}(a^-_{|j})) = C_\mu(\breve{u}(a^+_{|j})) = -\mu_j = -C_\mu(u(a_j)) \tag{16}$$

$$\breve{C}_\mu(\breve{u}(a_{i|j})) = C_\mu(\breve{u}(a^+_{i|j})) - C_\mu(\breve{u}(a^-_{i|j})) = \mu_i - \mu_j = C_\mu(u(a_i)) - C_\mu(u(a_j)) \tag{17}$$

$$\breve{C}_\mu(\breve{u}(a_{ij|})) = C_\mu(\breve{u}(a^+_{ij|})) - C_\mu(\breve{u}(a^-_{ij|})) = \mu_{ij} = C_\mu(u(a_{ij})) \tag{18}$$

$$\breve{C}_\mu(\breve{u}(a_{|ij})) = C_\mu(\breve{u}(a^+_{|ij})) - C_\mu(\breve{u}(a^-_{|ij})) = -\mu_{ij} = -C_\mu(u(a_{ij})) \tag{19}$$

Hence, due to the bipolarity of the scales used, to entirely determine the 2-additive capacity for the symmetric Choquet integral, we choose in this case to get some preferential information from the DM on trinary actions instead of binary actions. Furthermore the Equations 14 to 19 show that there is a strong link between binary and trinary actions. The preference information on $\mathcal{T}$ given by the DM is expressed by the two following relations:

- $\breve{P} = \{(x,y) \in \mathcal{T} \times \mathcal{T} :$ the DM strictly prefers $x$ to $y\}$,
- $\breve{I} = \{(x,y) \in \mathcal{T} \times \mathcal{T} :$ the DM is indifferent between $x$ and $y\}$.

**Definition 2.** *The ordinal information on* $\mathcal{T}$ *is the structure* $\{\breve{P}, \breve{I}\}$.

**Example 3.** *Given* $N = \{1,2,3\}$ *and* $\mathcal{T} = \{a_{0|0}; a_{1|}; a_{2|}; a_{3|}; a_{|1}; a_{|2}; a_{|3}; a_{1|2}; a_{1|3}; a_{2|1};$
$a_{2|3}; a_{3|1}; a_{3|2}; a_{12|}; a_{13|}; a_{|12}; a_{23|}; a_{|13}; a_{|23}\}$, *the following preferences are an ordinal information on* $\mathcal{T}$*:* $\breve{P} = \{(a_{2|1}, a_{3|2})\}$ *and* $\breve{I} = \{(a_{1|}, a_{2|})\}$

As in Section 2.2 we will suppose $\breve{P} \neq \emptyset$ in this section ("non-triviality" axiom) and will complete $\{\breve{P}, \breve{I}\}$ by the binary relation $\breve{M}$ on $\mathcal{T}$ defined as follows: for $(x,y) \in \{(a_{i|}, a_{0|0}), i \in N\} \cup \{(a_{0|0}, a_{|i}), i \in N\} \cup \{(a_{ij|}, a_{i|}), i,j \in N\} \cup \{(a_{|i}, a_{|ij}), i,j \in N\} \cup \{(a_{|i}, a_{i|j}), i,j \in N\} \cup \{(a_{i|j}, a_{|j}), i,j \in N\}$

$$x \breve{M} y \Leftrightarrow \text{ not } (x\ (\breve{P} \cup \breve{I})\ y)$$

The relation $\breve{M}$ models the natural monotonicity conditions for bicapacity (see Equation (13)): $\nu(\emptyset, \{i,j\}) \leq \nu(\emptyset, \{j\}) \leq \nu(\emptyset, \emptyset)$ and $\nu(\emptyset, \{j\}) \leq \nu(\{i\}, \{j\}) \leq \nu(\{i\}, \emptyset) \leq \nu(\{i,j\}, \emptyset)$

**Definition 3.** *An ordinal information* $\{\breve{P}, \breve{I}\}$ *on* $\mathcal{T}$ *is called a* $\mathcal{B}$*-ordinal information if it fulfills the following conditions: for all* $i,j,k,l \in N$

*1.*

$$\left.\begin{array}{c} a_{i|j}\ (\{\breve{P}\cup\breve{I}\})\ a_{k|} \\ or \\ a_{|k}\ (\{\breve{P}\cup\breve{I}\})\ a_{j|i} \\ or \\ a_{k|}\ (\{\breve{P}\cup\breve{I}\})\ a_{i|j} \\ or \\ a_{j|i}\ (\{\breve{P}\cup\breve{I}\})\ a_{|k} \\ or \\ a_{i|j}\ (\{\breve{P}\cup\breve{I}\})\ a_{kl|} \\ or \\ a_{|kl}\ (\{\breve{P}\cup\breve{I}\})\ a_{j|i} \\ or \\ a_{kl|}\ (\{\breve{P}\cup\breve{I}\})\ a_{i|j} \\ or \\ a_{j|i}\ (\{\breve{P}\cup\breve{I}\})\ a_{|kl} \end{array}\right\} \Rightarrow [a_{0|0}\ \breve{TC}\ a_{j|}] \tag{20}$$

*2.*

$$\left.\begin{array}{c} a_{i|j}\ (\{\breve{P}\cup\breve{I}\})\ a_{k|l} \\ or \\ a_{l|k}\ (\{\breve{P}\cup\breve{I}\})\ a_{j|i} \end{array}\right\} \Rightarrow [a_{l|}\ \breve{TC}\ a_{j|}\ \textit{and}\ a_{j|}\ \breve{TC}\ a_{l|}] \tag{21}$$

*where* $x\ \breve{TC}\ y \Leftrightarrow$ *there exists a path (not necessarily strict) of* $(\breve{P}\cup\ \breve{I}\cup\breve{M})$ *from* $x$ *to* $y$.

In fact, a $\mathcal{B}$-ordinal information is an ordinal information on $\mathcal{T}$ which can be treated as an ordinal information on $\mathcal{B}$. Generally, to compute a 2-additive capacity for symmetric Choquet integral, one asks only preferences on a positive or negative part of the scale. In our case, this type of preference information on $\mathcal{B}$ corresponds to a particular case of a $\mathcal{B}$-ordinal information.

### 3.2 Our Characterization Theorem

Given an ordinal information on $\mathcal{T}$, our aim is to find necessary and sufficient conditions for which there exists a 2-additive capacity such that $\{\breve{P},\breve{I}\}$ is representable by a symmetric Choquet integral i.e. for which there exists a 2-additive capacity $\mu$ satisfying:

$$(x,y)\in\breve{P}\Longrightarrow\breve{C}_\mu(\breve{u}(x))>\breve{C}_\mu(\breve{u}(y)) \tag{22}$$

$$(x,y)\in\breve{I}\Longrightarrow\breve{C}_\mu(\breve{u}(x))=\breve{C}_\mu(\breve{u}(y)) \tag{23}$$

Before to give our result, lets us show how to deduce an ordinal information on $\mathcal{B}$ from a $\mathcal{B}$-ordinal information on $\mathcal{T}$. Given $\{\breve{P},\breve{I}\}$ in this case, we construct the relations $P$ and $I$ on $\mathcal{B}$ as follows:

i. *Construction of* $P$:

Initially we set $P = \emptyset$. For all $i, j, k, l \in N$,

- if $a_{i|} \breve{P} a_{j|}$ or $a_{|i} \breve{P} a_{|j}$ then add $(a_i, a_j)$ in $P$;
- if $a_{i|} \breve{P} a_{kl|}$ or $a_{|kl} \breve{P} a_{|i}$ then add $(a_i, a_{kl})$ in $P$;
- if $a_{|i} \breve{P} a_{|kl}$ or $a_{kl|} \breve{P} a_{i|}$ then add $(a_{kl}, a_i)$ in $P$;
- if $a_{i|j} \breve{P} a_{k|}$ or $a_{|k} \breve{P} a_{j|i}$ then add $(a_i, a_k)$ in $P$;
- if $a_{k|} \breve{P} a_{i|j}$ or $a_{j|i} \breve{P} a_{|k}$ then add $(a_k, a_i)$ in $P$;
- if $a_{i|j} \breve{P} a_{kl|}$ or $a_{|kl} \breve{P} a_{j|i}$ then add $(a_i, a_{kl})$ in $P$;
- if $a_{kl|} \breve{P} a_{i|j}$ or $a_{j|i} \breve{P} a_{|kl}$ then add $(a_{kl}, a_i)$ in $P$;
- if $a_{i|j} \breve{P} a_{k|l}$ or $a_{l|k} \breve{P} a_{j|i}$ then add $(a_i, a_k)$ in $P$.

ii. *Construction of* $I$:

Initially we set $I = \emptyset$. For all $i, j, k, l \in N$,

- if $a_{i|} \breve{I} a_{j|}$ or $a_{|i} \breve{I} a_{|j}$ then add $(a_i, a_j)$ in $I$;
- if $a_{i|} \breve{I} a_{kl|}$ or $a_{|kl} \breve{I} a_{|i}$ then add $(a_i, a_{kl})$ in $I$;
- if $a_{|i} \breve{I} a_{|kl}$ or $a_{kl|} \breve{I} a_{i|}$ then add $(a_{kl}, a_i)$ in $I$;
- if $a_{i|j} \breve{I} a_{k|}$ or $a_{j|i} \breve{I} a_{|k}$ then add $(a_k, a_i)$ in $I$;
- if $a_{i|j} \breve{I} a_{kl|}$ or $a_{j|i} \breve{I} a_{|kl}$ then add $(a_i, a_{kl})$ in $I$;
- if $a_{i|j} \breve{I} a_{k|l}$ or $a_{l|k} \breve{I} a_{j|i}$ then add $(a_i, a_k)$ in $I$.

**Example 4.** *It is not difficult to see that* $\breve{P} = \{(a_{2|1}, a_{3|2})\}$ *and* $\breve{I} = \{(a_{1|}, a_{2|})\}$ *given in Example 3 is a* $\mathcal{B}$*-ordinal information on* $\mathcal{T}$*. Therefore, from* $\{\breve{P}, \breve{I}\}$ *we compute the following ordinal information on* $\mathcal{B}$*:* $P = \{(a_2, a_3)\}$ *and* $I = \{(a_1, a_2)\}$.

The following Theorem extends the characterization of ordinal information on $\mathcal{B}$ to a $\mathcal{B}$-ordinal information $\mathcal{T}$.

**Theorem 2.** *A* $\mathcal{B}$*-ordinal information on* $\mathcal{T}$*,* $\{\breve{P}, \breve{I}\}$*, is representable by a 2-additive symmetric Choquet integral if it fulfills these three conditions:*

1. *There is no strict cycle of* $(\breve{P} \cup \breve{I} \cup \breve{M})$
2. *There is no strict cycle of* $(P \cup I \cup M)$
3. *Any subset* $K$ *of* $N$ *such that* $|K| = 3$ *satisfies the MOPI property (using* $(P \cup I \cup M)$*).*

*Where* $\{P, I\}$ *is an ordinal information on* $\mathcal{B}$ *computed from* $\{\breve{P}, \breve{I}\}$ *as shown above.*

Knowing that the trinary actions are mainly useful to elicitate a 2-additive bi-capacity, this characterization gives a family of ordinal information on $\mathcal{T}$ compatible with a symmetric Choquet integral (a particular case of bipolar Choquet integral).

## References


1. Ah-Pine, J., Mayag, B., Rolland, A.: Identification of a 2-additive bi-capacity using mathematical programming. In: Perny, P., Pirlot, M., Tsoukiàs, A. (eds.) ADT 2013. LNCS, vol. 8176, pp. 15–29. Springer, Heidelberg (2013)
2. Bouyssou, D., Couceiro, M., Labreuche, C., Marichal, J.-L., Mayag, B.: Using choquet integral in machine learning: What can mcda bring? In: DA2PL 2012 Workshop: From Multiple Criteria Decision Aid to Preference Learning, Mons, Belgique (2012)
3. Grabisch, M.: $k$-order additive discrete fuzzy measures and their representation. Fuzzy Sets and Systems 92, 167–189 (1997)
4. Grabisch, M., Labreuche, C.: Fuzzy measures and integrals in MCDA. In: Figueira, J., Greco, S., Ehrgott, M. (eds.) Multiple Criteria Decision Analysis: State of the Art Surveys, pp. 565–608. Springer (2005)
5. Grabisch, M., Labreuche, C.: The symmetric and asymmetric Choquet integrals on finite spaces for decision making. Statistical Papers 43, 37–52 (2002)
6. Grabisch, M., Labreuche, C.: A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. 4OR 6, 1–44 (2008)
7. Grabisch, M., Labreuche, C., Vansnick, J.-C.: On the extension of pseudo-Boolean functions for the aggregation of interacting bipolar criteria. Eur. J. of Operational Research 148, 28–47 (2003)
8. Grabisch, M., Labreuche, C.: Bi-capacities-I: definition, Möbius transform and interaction. Fuzzy Sets and Systems 151(2), 211–236 (2005)
9. Grabisch, M., Labreuche, C.: Bi-capacities-II: the Choquet integral. Fuzzy Sets and Systems 151(2), 237–259 (2005)
10. Grabisch, M., Labreuche, C.: Fuzzy measures and integrals in MCDA. In: Multiple Criteria Decision Analysis: State of the Art Surveys. Int. Series in Op. Res. & Manag. Sci., vol. 78, pp. 563–604. Springer, New York (2005)
11. Labreuche, C., Grabisch, M.: The Choquet integral for the aggregation of interval scales in multicriteria decision making. Fuzzy Sets and Systems 137, 11–26 (2003)
12. Mayag, B., Grabisch, M., Labreuche, C.: A characterization of the 2-additive Choquet integral through cardinal information. Fuzzy Sets and Systems 184(1), 84–105 (2011)
13. Mayag, B., Grabisch, M., Labreuche, C.: A representation of preferences by the Choquet integral with respect to a 2-additive capacity. Theory and Decision 71(3), 297–324 (2011)
14. Mayag, B., Rolland, A., Ah-Pine, J.: Elicitation of a 2-additive bi-capacity through cardinal information on trinary actions. In: Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R.R. (eds.) IPMU 2012, Part IV. CCIS, vol. 300, pp. 238–247. Springer, Heidelberg (2012)
15. Murofushi, T., Soneda, S.: Techniques for reading fuzzy measures (III): interaction index. In: 9th Fuzzy System Symposium, Sapporo, Japan, pp. 693–696 (May 1993) (in Japanese)
16. Shapley, L.S.: A value for $n$-person games. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games. Annals of Mathematics Studies, vol. II(28), pp. 307–317. Princeton University Press (1953)
17. Simon, H.: Rational choice and the structure of the environment. Psychological Review 63(2), 129–138 (1956)
18. Slovic, P., Finucane, M., Peters, E., MacGregor, D.G.: The affect heuristic. In: Gilovitch, T., Griffin, D., Kahneman, D. (eds.) Heuristics and biases: the Psychology of Intuitive Judgment, pp. 397–420. Cambridge University Press (2002)
19. Tversky, A., Kahneman, D.: Advances in prospect theory: cumulative representation of uncertainty. J. of Risk and Uncertainty 5, 297–323 (1992)

# Choquet Integral on Multisets

Yasuo Narukawa[1,2] and Vicenç Torra[3]

[1] Toho Gakuen,
3-1-10, Naka, Kunitachi, Tokyo, 186-0004, Japan
[2] Department of Computational Intelligence and Systems Science
Tokyo Institute of Technology 4259 Nagatuta, Midori-ku, Yokohama 226-8502, Japan
[3] IIIA - CSIC, Campus UAB s/n,
08193 Bellaterra, Catalonia, Spain
narukawa@d4.dion.ne.jp, vtorra@iiia.csic.es

**Abstract.** Fuzzy measures on multisets are studied in this paper. We show that a class of multisets on a finite space can be represented as a subset of positive integers. Comonotonicity for multisets are defined. We show that a fuzzy measure on multisets with some comonotonicity condition can be represented by a generalized fuzzy integral.

**Keywords:** Fuzzy measure, multiset, Choquet integral, Sugeno integral, Generalized fuzzy integral.

## 1 Introduction

Multisets or bags are a generalization of sets in which multi-occurrences of elements are allowed. They have been used in the areas of mathematics and computer science [12], among other disciplines.

Standard set operations as union and intersection have been defined for multisets. Other operations have been defined as well. In this paper we focus on fuzzy measures and integrals on multisets. Some preliminary results on fuzzy measures on multisets were presented in [17,10].

Fuzzy measures, also known as non-additive measures, are monotonic set functions on a reference set. They generalize additive measures (probabilities) not requiring the measure to be additive and replacing this axiom by the one of monotonicity. Fuzzy integrals have been defined which integrate a function with respect to a non-additive measure. The most well known fuzzy integrals are the Sugeno integral [13,14] and the Choquet integral [3]. See [16] for a state-of-the-art description of the field.

In this paper we focus on fuzzy measures and Choquet integral on multisets. We show that a class of multisets can be represented as subsets of positive integers. We also define comonotonicity for multisets, and we show that a fuzzy measure on multisets, with some comonotonicity condition, can be represented by a Choquet integral.

The structure of the paper is as follows. In Section 2 we review some definitions that are needed in the rest of the paper. In Section 3 we focus on the representation of integrals on finite multisets. In Section 4, we present a way to extend a fuzzy measure on the class of measurable sets to the class of multisets. We finish the paper with some concluding remarks on Section 5.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 276–283, 2014.

## 2 Preliminaries

This section is divided in two parts. First we review the definition of Choquet and Sugeno integrals, and then some definitions related to multisets.

### 2.1 Choquet Integral

Here we present the definition of fuzzy measures and of the Choquet integral. The Choquet integral integrates a function with respect to a fuzzy measure.

**Definition 1.** *Let $X$ be a universal set and $\mathscr{X}$ be a $\sigma$-algebra of $2^X$. Then, $(X,\mathscr{X})$ is called a fuzzy measurable space. We say that a function $f : X \to \mathbb{R}^+$ is $\mathscr{X}$-measurable if $\{x|f(x) \geq a\} \in \mathscr{X}$ for all $a$.*

**Definition 2.** *[4] Let $f$ and $g$ be $\mathscr{X}$-measurable functions on $X$; then, we say that $f$ and $g$ are comonotonic if*

$$f(x) < f(y) \Rightarrow g(x) \leq g(y)$$

*for $x, y \in X$.*

**Definition 3.** *[14] Let $(X,\mathscr{X})$ be a fuzzy measurable space; then, a fuzzy measure $\mu$ on $(X,\mathscr{X})$ is a real valued set function, $\mu : \mathscr{X} \longrightarrow \mathbb{R}^+$ with the following properties.*

(i) $\mu(\emptyset) = 0$, $\mu(X) = k$ *where* $k \in (0,\infty)$.
(ii) $\mu(A) \leq \mu(B)$ *whenever* $A \subset B$, $A, B \in \mathscr{X}$.

*A triplet $(X,\mathscr{X},\mu)$ is said to be a fuzzy measure space.*

**Definition 4.** *[3,8] Let $(X,\mathscr{X},\mu)$ be a fuzzy measure space and let $f$ be a non-negative $\mathscr{X}$-measurable function; then, the Choquet integral of $f$ with respect to $\mu$ is defined by*

$$(C)\int_X f d\mu := \int_0^\infty \mu_f(r)dr,$$

*where $\mu_f(r) = \mu(\{x|f(x) \geq r\})$.*

Let $X = \{x_1, x_2, \ldots, x_n\}$ and $f(x_1) < f(x_2) < \cdots < f(x_n)$. Then the Choquet integral of $f$ is expressed as

$$(C)\int_X f d\mu = \sum_{k=1}^{n} (f(x_k) - f(x_{k-1}))\mu(\{x_k, \ldots, x_n\})$$

where $f(x_0) = 0$.

Let $f, g$ be comonotonic measurable functions. Then, since for all $a, b > 0$ either $\{x|f(x) \geq a\} \subset \{x|g(x) \geq b\}$ or $\{x|f(x) \geq a\} \supset \{x|g(x) \geq b\}$, the following theorem can be proved.

**Theorem 1.** *[4] Let $(X,\mathscr{X},\mu)$ be a fuzzy measure space. Then, for comonotonic measurable functions $f$, and $g$, we have*

$$(C)\int_X (f+g)d\mu = (C)\int_X f d\mu + (C)\int_X g d\mu.$$

We call this property the comonotonic additivity of a Choquet integral.

### 2.2 Multisets

Let $X$ be a universal set. Then, a multiset $M$ of $X$ is characterized by the count function $C_M : X \to N := \{0,1,2,\dots\}$. Here, $C_M$ corresponds to the number of occurrences of the object $x \in X$.

We denote by $\mathscr{M}(X)$ the class of multisets of $X$.

**Example 1.** *Let $X := \{a,b,c\}$ and $M := (a,a,a,b,b)$, that is, $C_M(a) = 3$, $C_M(b) = 2$, $C_M(c) = 0$.*

We can express $M$ in Example 1 either as $M = (3/a, 2/b)$ or as $M = ((a,3),(b,2))$.

**Definition 5.** *Let $M,N \in \mathscr{M}(X)$. Then, we define:*

- *the inclusion of multisets by*

$$M \subset N \Leftrightarrow C_M(x) \leq C_N(x)$$

  *for all $x \in X$;*
- *the equality of multisets $M = N$ by*

$$C_M(x) = C_N(x).$$

Let $M \in \mathscr{M}(X)$. Then $\mathscr{P}(M)$ denotes the class of subsets of multiset $M$, that is,

$$\mathscr{P}(M) := (N | N \subset M, N \in \mathscr{M}(X)\}.$$

**Proposition 1.** *Let $|X| = n$ and $M \in \mathscr{M}(X)$. If $M = ((a_i,k_i)|i = 1,2,\dots n\}$, then*

$$|\mathscr{P}(M)| = \Pi_{i=1}^{n}(k_i + 1).$$

**Example 2.** *Let $M = (a,a,a,b,b)$, $M_0 = \emptyset$, $M_1 = (a)$, $M_2 = (a,a)$, $M_3 = (a,a,a)$, $M_4 = (a,a,a,b)$, $M_5 = (a,a,a,b,b)$, $M_6 = (a,a,b)$, $M_7 = (a,a,b,b)$, $M_8 = (a,b)$, $M_9 = (a,b,b)$, $M_{10} = (b)$, $M_{11} = (b,b)$.*

*Then $\mathscr{P}(M) = \{M_i | i = 0,1,2,\dots,11\}$.*

**Definition 6.** *Let $A,B \in \mathscr{M}(X)$, and let $\alpha \in N$. Then, we define some binary operations on $\mathscr{M}(X)$. Definitions include union, intersection, addition of two multisets and multiplication of a multiset by an integer. In these definitions, $\vee$ corresponds to the maximum and $\wedge$ to the minimum.*

(i) $C_{A\cup B}(x) = C_A(x) \vee C_B(x)$
(ii) $C_{A\cap B}(x) = C_A(x) \wedge C_B(x)$
(iii) $C_{A+B}(x) = C_A(x) + C_B(x)$
(iv) $C_{\alpha A}(x) = \alpha C_B(x)$

*where $x \in X$ and $C_A$ is the count function of $A$.*

**Proposition 2.** *Let $A,B \in \mathscr{M}(X)$. We have*

$$A \cap B \subset A \cup B \subset A + B$$

**Example 3.** *Let $X := \{a,b,c\}$ and $A := (a,a,b)$, $B := (a,b,b,c)$. Then we have*

(i) $A \cup B = (a,a,b,b,c)$
(ii) $A \cap B = (a,b)$
(iii) $A + B = (a,a,a,b,b,b,c)$

## 3 Repesentation of Integral on Finite Multisets

Let $X$ be a finite universal set and $|X| = n$.

Let $P$ be the set of prime numbers, that is, $P := \{2,3,5,7,\ldots,\}$.

Since $X$ is a finite set, there exists a one to one mapping $\varphi_X$ from $X$ to a subset of $P$, that is, $\varphi_X : X \to \{p_1, p_2, \ldots, p_n\}$.

Let $M \in \mathscr{M}(X)$. We have an induced one to one mapping $\Phi_X$ from $\mathscr{M}(X)$ to a subset $S$ of natural numbers by $\Phi_X(M) := \Pi_{i=1}^{n} \varphi_X(x_i)^{C_M(x_i)}$. In this case we say that $\Phi_X(M)$ is a natural number representation of a multiset $M$.

**Example 4.** *Let $X := \{a,b,c\}$ and $\varphi_X(a) = 2, \varphi_X(b) = 3, \varphi_X(c) = 5$. Then $\Phi_X(A) := 2^{C_A(a)}3^{C_A(b)}5^{C_A(c)}$ for $A \in \mathscr{M}(X)$.*

*In fact, if $A := (a,a,b,c)$, then $\Phi_X(A) = 2^2 \cdot 3 \cdot 5 = 60$.*

Let $M \in \mathscr{M}(X)$. We have

$$\Phi_X(\mathscr{P}(M)) := \{\Pi_{i=1}^{n} \varphi_X(x_i)^{C_A(x_i)} | A \in \mathscr{P}(M)\}.$$

**Proposition 3.** *Let $M \in \mathscr{M}(X)$; then, $\Phi_X(\mathscr{P}(M))$ is a set of divisors of $\Phi_X(M)$.*

**Example 5.** *Let $X := \{a,b,c\}$ and $\varphi_X(a) = 2, \varphi_X(b) = 3, \varphi_X(c) = 5$.*

*Let $M := (a,a,a,b,b,c)$.*

*Then $\Phi_X(M) := 2^3 3^2 5^1 = 120$, and*

$$\Phi_X(\mathscr{P}(M)) := \{1,2,3,4,5,6,8,10,12,15,24,30,40,60,120\}.$$

**Theorem 2.** *Let $\mathscr{N} \subset \mathscr{M}(X)$.*

*For $f : \mathscr{N} \to \mathbb{R}^+$ and $\mu$ be a fuzzy measure on $2^{\mathscr{N}}$ there exist a subset $N_0$ of a set of non negative integers $\mathbb{N}$, $f_{\mathscr{N}} : N_0 \to \mathbb{R}^+$ and $\mu_{\mathscr{N}}$ : a fuzzy measure on $2^{N_0}$ such that*

$$(C)\int_{\mathscr{N}} f d\mu = (C)\int_{N_0} f_{\mathscr{N}} d\mu_{\mathscr{N}}.$$

*Conversely the Choquet integral on the set of positive integer can be regarded as the Choquet integral on multisets.*

**Remark.** Since a multiset on a finite universal space can be regarded as a positive integer, a class of multisets can be regarded as a subset of the positive integers, and a function on a class of multisets as a function on a subsets of positive integers. This shows that to study a function and the integral on the class of multisets, which seems complicated, is identical to the study of functions and integrals on subsets of positive integers as long as the universal set is finite.

As a special case of this fact, a fuzzy measure and the integral on the discrete space is identical with the ones on finite subsets of prime numbers.

From this point of view, it is natural to study the fuzzy integral on multisets if the universal space is finite or countable.

If the universal space is not countable, the representation above is impossible.

**Example 6.** *Let $X := \{a,b\}$ and $\mathscr{N} := \{(a),(a,a),(a,b)\}$. Then, define a function $f$ on $\mathscr{N}$ by*

$$f((a)) = 9, f((a,a)) = 3, f((a,b)) = 5,$$

*and a fuzzy measure on $2^{\mathscr{N}}$ by*
$\mu(\{(a)\}) = 1$, $\mu(\{(a,a)\}) = 1.8$, $\mu(\{(a,b)\}) = 1.5$, $\mu(\{(a),(a,a)\}) = 2.5$,
$\mu(\{(a),(a,b)\}) = 2.2$, $\mu(\{(a,a),(a,b)\}) = 3$. $\mu(\mathscr{N}) = 4$
*Then we have*

$$\begin{aligned}(C)\int_{\mathscr{N}} f d\mu &= (f((a)) - f((a,b)))\mu(\{(a)\}) + (f((a,b)) - f((a,a)))\mu(\{(a),(a,b)\}) \\ &\quad + f((a,a))\mu(\mathscr{N}) \\ &= 4\times 1 + 2\times 2.5 + 3\times 4 = 21\end{aligned}$$

*Using a natural number representation, we have $N_0 = \{2,4,6\}$,*
$f_{\mathscr{N}}(2) = 9, f_{\mathscr{N}}(4) = 3, f_{\mathscr{N}}(6) = 5$, *and*
$\mu_{\mathscr{N}}(\{2\}) = 1$, $\mu_{\mathscr{N}}(\{4\}) = 2$, $\mu_{\mathscr{N}}(\{6\}) = 1.5$, $\mu_{\mathscr{N}}(\{2,4\}) = 2.5$, $\mu_{\mathscr{N}}(\{2,6\}) = 2.2$,
$\mu_{\mathscr{N}}(\{4,6\}) = 3$. $\mu_{\mathscr{N}}(N_0) = 4$.

The example above can be regarded under the next scenario.

**Scenario 1.** *The supermarket sells several packed fruits. Let us consider the following ones.*

- $(a)$*: one orange*
- $(a,a)$*:one pack of two oranges*
- $(a,b)$*: one pack of one orange and one apple.*

*Note that a single apple is not sold there.*

*Let us consider that a customer, Tom, wants to buy 20 oranges and 5 apples.*

*Then, the fuzzy measure $\mu(*)$ is the price of packed fruits, which has discount for a specific combination. In addition, the function $f(A)$ corresponds to the number of packed fruits A that the customer buys.*

*Then the Choquet integral of $f$ with respect to $\mu$ is the total price Tom should pay.*

In such scenario, the problem is to minimize the total price. In Example 6 we had that the Choquet integral is 21. Alternatively, if we define the function $g$ on $\mathscr{N}$ by

$$g((a)) = 1, g((a,a)) = 7, g((a,b)) = 5,$$

then the total is

$$\begin{aligned}(C)\int_{\mathscr{N}} g d\mu &= (g((a,a)) - g((a,b)))\mu(\{(a,a)\}) + (g((a,b)) - g((a)))\mu(\{(a,a),(a,b)\}) \\ &\quad + f((a))\mu(\mathscr{N}) \\ &= 2\times 1.8 + 4\times 3 + 1\times 4 = 19.6\end{aligned}$$

Therefore $(C)\int g d\mu < (C)\int f d\mu$, that is, $g$ is better selection of packs than $f$.

For this scenario, for established prices $\mu$, when Tom wants to buy the multiset $M \in \mathscr{M}$, his best selection with respect to the final price would be the function $f : \mathscr{N} \to \mathbb{R}^+$ such that

$$M \leq \sum_{A \subset \mathscr{N}} f(A) \cdot 1_A \tag{1}$$

and with a minimal Choquet integral. Formally:

Let $F = \{f | M \leq \sum_{A \subset \mathscr{N}} f(A) \cdot 1_A\}$, Tom's best selection is

$$\arg\min_{f \in F} (C) \int f d\mu.$$

## 4 Distorted Measure on Multisets and Integral

Let $X$ be finite. We can define a fuzzy measure $\mu$ on $2^X$. Let $M$ be a multiset on $X$, then we can represent a count function $C_M$ by $C_M := \sum_{i=1}^{n} a_i 1_{A_i}$, with $a_i > 0$ and $A_1 \supsetneq A_2 \supsetneq \ldots A_n$, $A_i \in \mathscr{X}$. We call this representation the comonotonic additive representation of $C_M$. Note that the comonotonic additive representation of a $M$ is unique.

Then we can define an extension $\bar{\mu}$ of fuzzy measure $\mu$ to a multiset $M$ by

$$\bar{\mu}(M) := \sum_{i=1}^{n} a_i \cdot \mu(A_i)$$

with $a_i > 0$ and $A_1 \supsetneq A_2 \supsetneq \ldots A_n$, $A_i \in \mathscr{X}$.

We say that $\bar{\mu}$ is a comonotonic extension of $\mu$.

Since

$$\bar{\mu}(M) = (C) \int_X C_M d\mu,$$

we have the next proposition.

**Proposition 4.** *[10] Let $M,N \in \mathscr{M}(X)$, and let $\mu$ be a fuzzy measure on $2^X$. If $M$ and $N$ are comonotonic, then $\bar{\mu}(M+N) = \bar{\mu}(M) + \bar{\mu}(N)$.*

Let $g : R^+ \to R^+$ be monotone and $g(0) = 0$. We can define a fuzzy measure $\eta$ on $\mathscr{N} \subset M(X)$ by

$$\eta(A) = g(\sum_{M \in A} \bar{\mu}(M)) \tag{2}$$

for $A \in \mathscr{N}$.

We say that $\eta$ is a distorted measure generated by $\mu$ and $g$.

**Corollary 1.** *Let $\mathscr{N} \subset \mathscr{M}(X)$. Then, for a function $f : \mathscr{N} \to \mathbb{R}^+$ and a distorted measure $\eta$ generated by $\mu$ and $g$ on $2^{\mathscr{N}}$ there exist a subset $N_0$ of a set of non negative integers $\mathbb{N}$, $f_{\mathscr{N}} : N_0 \to \mathbb{R}^+$ and $\mu_{\mathscr{N}}$ : a distorted measure on $2^{N_0}$ such that*

$$(C)\int_{\mathscr{N}} f d\eta = (C)\int_{N_0} f_{\mathscr{N}} d\mu_{\mathscr{N}}.$$

**Example 7.** *Let* $X := \{a,b,c\}$. *Then, let us define the fuzzy measure* $\mu$ *on* $2^X$ *by*
$\mu(\{a\}) = 1$, $\mu(\{b\}) = 1.5$, $\mu(\{c\}) = 2$, $\mu(\{a,b\}) = 2$, $\mu(\{b,c\}) = 3$,
$\mu(\{a,c\}) = 2.5$, $\mu(\{a,b,c\}) = 4$.
*Let* $M := \{a,a,a,a,b,b,c\}$.
*We have a comonotonic additive representation of* $C_M$ *by*
$C_M = 1_{\{a,b,c\}} + 1_{\{a,b\}} + 2 \times 1_{\{a\}}$.
*Then we have the comonotonic extension of* $\mu$ *by*
$\bar{\mu}(M) = \mu(\{a,b,c\}) + \mu(\{a,b\}) + 2 \times \mu(a) = 4 + 2 + 2 \times 1 = 8$.
*Let* $\mathcal{N} := \{(a),(c),(a,a,a,b),M\}$ *and* $g(x) = x^2$.
*Then we have a distorted measure* $\eta$ *by* $\eta(A) = (\sum_{B \in A} \bar{\mu}(B))^2$.
*Using a natural number representation,* $N_0 = \{2,5,24,720\}$.
*Define a function* $f : \mathcal{N} \to R^+$ *by*
$f((a)) = 0$, $f((c)) = 4$, $f((a,a,a,b)) = 5$, $f(M) = 2$.
*That is,* $f_{\mathcal{N}}(2) = 0$, $f_{\mathcal{N}}(5) = 4$, $f_{\mathcal{N}}(24) = 5$, $f_{\mathcal{N}}(720) = 2$,
$\mu_{\mathcal{N}}(\{24\}) = \eta((a,a,a,b)) = (\bar{\mu}((a,a,a,b)))^2 = (2\mu((a)) + \mu((a,b)))^2 = 16$
$\mu_{\mathcal{N}}(\{5,24\}) = \eta((c),(a,a,a,b)) = (\bar{\mu}((c)) + \bar{\mu}(a,a,a,b)))^2 = (2+4)^2 = 36$

$$\begin{aligned}\mu_{\mathcal{N}}(\{5,24,720\}) &= \eta((c),(a,a,a,b),M) \\ &= (\bar{\mu}(\{c\}) + \bar{\mu}((a,a,a,b)) + \bar{\mu}(M))^2 \\ &= (2+4+8)^2 = 196\end{aligned}$$

$$\begin{aligned}\mu_{\mathcal{N}}(\{2,5,24,720\}) &= \eta((a),(c),(a,a,a,b),M) \\ &= (\bar{\mu}((a)) + \bar{\mu}((c)) + \bar{\mu}((a,a,a,b)) + \bar{\mu}(M))^2 \\ &= (1+2+4+8)^2 = 225\end{aligned}$$

*We have*

$$\begin{aligned}&(C)\int_{\mathcal{N}} f d\eta = (C)\int_{N_0} f_{\mathcal{N}} d\mu_{\mathcal{N}} \\ &= (f_{\mathcal{N}}(24) - f_{\mathcal{N}}(5))\mu_{\mathcal{N}}(\{24\}) + (f_{\mathcal{N}}(5) - f_{\mathcal{N}}(720))\mu_{\mathcal{N}}(\{5,24\}) \\ &+ (f_{\mathcal{N}}(720) - f_{\mathcal{N}}(2))\mu_{\mathcal{N}}(\{5,24,720\}) + f_{\mathcal{N}}(2)\mu_{\mathcal{N}}(\{2,5,24,720\}) \\ &= 1 \times 16 + 2 \times 36 + 2 \times 196 + 0 \times 225 = 480\end{aligned}$$

**Remark.** We use $\Sigma$ to define a distorted measure $\eta$ in Equation 2. This can be extended using a binary operation with associativity [7,1]. Moreover to define a comonotonic extension $\mu$, we can use the generalized fuzzy integral [9] instead of Choquet integral. We can chose the suitable operation or integral instead of Choquet integral with respect to a distorted measure.

## 5 Conclusion

In this paper we have studied fuzzy measures and Choquet integral for multisets on a finite universal space. We have shown in Theorem 2 that Choquet integrals for multisets

can be seen as a Choquet integral on a set of positive numbers. We have also shown an scenario where this type of measures and integrals can be applied. Finally, we have presented some results about distorted measures on multisets. We expect to have deeper results in the future using number theory. In particular, we will focus on the extension of the results to a generalized fuzzy integral, including the Sugeno integral.

**Acknowledgements.** Partial support by the Spanish MEC projects ARES (CONSOLIDER INGENIO 2010 CSD2007-00004) and COPRIVACY (TIN2011-27076-C03-03) is acknowledged.

## References

1. Alsina, C., Frank, M.J., Schweizer, B.: Associative Functions: Triangular Norms and Copulas. World Scientific (2006)
2. Benvenuti, P., Mesiar, R., Vivona, D.: Monotone set functions-based integrals. In: Pap, E. (ed.) Handbook of Measure Theory, pp. 1329–1379. Elsevier, Amsterdam (2002)
3. Choquet, G.: Theory of capacities. Ann. Inst. Fourier 5, 131–295 (1953/1954)
4. Dellacherie, C.: Quelques commentaires sur les prolongements de capacités, Séminaire de Probabilités 1969/1970, Strasbourg. Lecture Notes in Mathematics, vol. 191, pp. 77–81 (1971)
5. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer Academic Publishers, Dordrecht (2000)
6. Klement, E.P., Mesiar, R., Pap, E.: Integration with respect to decomposable measures, based on a conditionally distributive semiring on the unit interval. Int. J. of Unc., Fuzziness and Knowledge Based Systems 8(6), 701–717 (2000)
7. Ling, C.H.: Representation of associative functions. Publ. Math. Debrecen 12, 189–212 (1965)
8. Murofushi, T., Sugeno, M.: An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure. Fuzzy Sets and Systems 29, 201–227 (1989)
9. Narukawa, Y., Torra, V.: Multidimensional generalized fuzzy integral. Fuzzy Sets and Systems 160, 802–815 (2009)
10. Narukawa, Y., Stokes, K., Torra, V.: Fuzzy Measures and Comonotonicity on Multisets. In: Torra, V., Narakawa, Y., Yin, J., Long, J. (eds.) MDAI 2011. LNCS, vol. 6820, pp. 20–30. Springer, Heidelberg (2011)
11. Ralescu, D., Adams, G.: The fuzzy integral. J. Math. Anal. Appl. 75, 562–570 (1980)
12. Raman, R., Raman, V., Rao, S.S.: Succint indexable dictionaries with applications to encoding k-ary trees and multisets. In: Proc. of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2002) (2002)
13. Sugeno, M.: Fuzzy measures and fuzzy integrals. Trans. of the Soc. of Instrument and Control Engineers 8(2) (1972) (in Japanese)
14. Sugeno, M.: Theory of fuzzy integrals and its applications, Ph. D. Dissertation, Tokyo Institute of Technology (1974)
15. Sugeno, M., Murofushi, T.: Pseudo-additive measures and integrals. J. Math. Anal. Appl. 122, 197–222 (1987)
16. Torra, V., Narukawa, Y., Sugeno, M.: Non-additive measures: theory and applications. Springer (2014)
17. Torra, V., Stokes, K., Narukawa, Y.: An extension of fuzzy measures to multisets and its relation to distorted probabilities. IEEE Trans. on Fuzzy Systems 20(6), 1032–1045 (2012)

# Rescaling for Evaluations Using Inclusion-Exclusion Integral

Aoi Honda[1], Ryoji Fukuda[2], and Jun Okamoto[3]

[1] Kyushu Institute of Technology, Department of Systems Design and Informatics, 680-2 Kawazu, Iizuka, Fukuoka 820-8502, Japan
aoi@ces.kyutech.ac.jp
[2] Oita University, Faculty of Engineering, 700 Dan-noharu 870-1192 Oita, Japan
rfukuda@oita-u.ac.jp
[3] NTT Network Technology Laboratories, NTT Corporation
okamoto.jun@lab.ntt.co.jp

**Abstract.** On multivariate analyses generally distributions of explanatory variable have deviation depending on each unique quality, and eliminating deviation often beneficially effective. We propose two algorithms for rescaling of raw data and verify the validity of them using real reliable big data.

**Keywords:** monotone measure, Choquet integral, multivariate analyses, inclusion-exclusion integral.

## 1 Introduction

Recently, data is becoming important for us and as the data itself grows dramatically, the storage and data management solutions that we correct become more critical. With such data we must attempt multivariate analyses and find the best way to analyze each data for the sake of getting various information from it. Our aim is, roughly speaking, to predict or estimate the value of a response variable from the known values explanatory variables or predictors, so that we have to determine models which yields the value of response variable by several explanatory variables. The most widely used model is the multiple linear regression model which specifies a response variable as a linear combination of independent explanatory variables. On the other hand, other models are proposed, for example, the Choquet integral model using monotone measures can express interactions among explanatory variables. These data, especially explanatory variables have deviation in the distribution depending on each special quality and several authors have been also investigating from this point of view[1]. In this study, we propose two algorithms for rescaling of raw data to eliminate deviation and using real big data, and verify the validity of them.

## 2 Preliminaries

Throughout this paper, the whole set is denoted by $N := \{1, 2, \ldots, n\}$ and $2^N$ denotes the power set of $N$. The minimum and the maximum operation in $2^N$

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 284–293, 2014.

are denoted by $\wedge$ and $\vee$, respectively. For a set $A$, the number of elements of $A$ is denoted by $|A|$.

### 2.1 Mathematical Preparation

**Definition 1 (Monotone Measure[2]).** *A set function $v : 2^N \to [0,1]$ is called a monotone measure if it satisfies*
1. $v(\emptyset) = 0, v(N) < +\infty$, *and*
2. $v(A) \leq v(B)$ *whenever* $A \subset B, A, B \in 2^N$.

**Definition 2 (Choquet Integral[3]).** *Let $v$ be a monotone measure defined on $2^N$, and $f$ a non-negative function on $N$. The Choquet integral of $f$ with respect to $v$ is defined by*

$$(C)\int f\ dv := \sum_{i=1}^{n}\big(f(\sigma(i)) - f(\sigma(i-1))\big)v(\{\sigma(i),\ldots,\sigma(n)\}),$$

*where $\sigma$ is a permutation on $N$ such that $f(\sigma(1)) \leq \cdots \leq f(\sigma(n))$ and $f(\sigma(0)) := 0$.*

**Definition 3 (Möbius transform).** *Let $v$ be a set function on $2^N$. The Möbius transform of $v$, denoted by $m^v$, is defined by*

$$m^v(A) := \sum_{B\subset A}(-1)^{|A\setminus B|}v(B)$$

*for any $A \in 2^N$. And $v$ and $m^v$ are one-to-one correspondence with*

$$v(A) = \sum_{B\subset A} m^v(B)$$

*for any $A \in 2^N$.*

**Proposition 1 ([4]).** *The Choquet integral of $f$ with respect to $v$ is represented with Möbius transform of $v$ by*

$$(C)\int f dv = \sum_{A\subset N}\left(\bigwedge_{i\in A} f(i)\right) m^v(A).$$

**Definition 4 ($k$-additive Measure[5]).** *Let $v$ be a monotone measure on $2^N$ and $k$ a positive integer. A monotone measure $v$ which satisfies $m^v(A) = 0$ whenever $|A| > k$ is called $k$-additive measure.*

$k$-additive measure admits an interpretation that this measure has interactions only among $k$ elements. The proposition below shows that the Choquet integral is represented as a linear combination of the Möbius transforms of $v$.

The inclusion-exclusion integral (IE integral) is defined by as follows.

**Definition 5 (Inclusion-Exclusion Integral[6, 7]).** *Let $v$ be a positive monotone measure defined on $2^N$, $\otimes : \cup_{k=1}^{n} \mathbb{R}^k \to \mathbb{R}$ be an $n$-ary operation on $\mathbb{R}$ and let $f$ be a function $N \to \mathbb{R}$. The inclusion-exclusion integral of $f$ with respect to $v$ is defined by*

$$\otimes \int f \; dv := \sum_{A \subset N} \left\{ \sum_{B \supset A} \left( (-1)^{|B \setminus A|} \bigotimes_{i \in B} f(i) \right) \right\} v(A),$$

*where $\otimes_{i \in B} w_i = \otimes \{ w_i \mid i \in B \}, B \subset N$.*

Remark that in [6, 7], $\otimes$ is assumed $t$-norm. But, in this paper we generalize it to general $n$-ary operation. Product-like operations such as $t$-norms are suitable as the $n$-ary operation of the inclusion-exclusion integral.

The IE integral can be represented by another form(Cf. [8–10]).

**Proposition 2.** *The inclusion-exclusion integral is represented with Möbius transform of $v$ by*

$$\otimes \int f \; dv = \sum_{A \subset N} \left( \bigotimes_{i \in A} f(i) \right) m^v(A).$$

By Propositions 1 and 2, we can regard the Choquet integral as an IE integral, then we obtain IE integral representation of the Choquet integral:

$$(C) \int f dv = \wedge \int f dv = \sum_{A \subset N} \left\{ \sum_{B \supset A} \left( (-1)^{|B \setminus A|} \bigwedge_{i \in B} f(i) \right) \right\} v(A),$$

which enables us to calculate the Choquet integral without rearrangement of $f$ or Möbius transform of $v$.

**Definition 6 ($T$-norm [11–13] etc.).** *If a binary operation $\otimes : [0,1] \times [0,1] \to [0,1]$ satisfies*

1. *$0 \otimes 0 = 0, x \otimes 1 = x$ for any $x > 0$,*
2. *$x \leq y$ implies $x \otimes z \leq y \otimes z$,*
3. *$x \otimes y = y \otimes x$ and*
4. *$x \otimes (y \otimes z)) = ((x \otimes y) \otimes z)$,*

*then $\otimes$ is called a $T$-norm.*

By *4* in Definition 6, $t$-norm $\otimes$ is extended to $n$-ary operation $[0,1]^n \to [0,1]$.

**Definition 7 (Dombi's $t$-norm [14]).** *Dombi's $t$-norm $\odot : [0,1]^2 \to [0,1]$ is defined by*

$$x \odot y := \frac{1}{1 + \left( \left( \frac{1}{x} - 1 \right)^{\lambda} + \left( \frac{1}{y} - 1 \right)^{\lambda} \right)^{\frac{1}{\lambda}}}, \quad \lambda > 0.$$

### 2.2 Regression Analyses Models

In the latter half of Section 2, we describe regression analyses and introduce several models needed later. Let $n, m$ be natural numbers. In this study, we consider $m$ data sets of response variable and $n$ explanatory variables, or predictors. One data set is consisted of explanatory variables $(x_1^j, x_2^j, \ldots, x_i^j, \ldots, x_n^j)$ and response variable $y^j$. Our objective is to select the most similar identification model to response using explanatory variables, in other words, to find a functions $F$ which explains the relation between $(x_1^j, x_2^j, \ldots, x_i^j, \ldots, x_n^j)$ and $y^j$ by $y^j = F(x_1^j, x_2^j, \ldots, x_i^j, \ldots, x_n^j)$.

**Table 1.** Data sets

| Data No. | $x_1$ | $x_2$ | $\cdots$ | $x_i$ | $\cdots$ | $x_n$ | $y$ |
|---|---|---|---|---|---|---|---|
| 1 | $x_1^1$ | $x_2^1$ | $\cdots$ | $x_i^1$ | $\cdots$ | $x_n^1$ | $y^1$ |
| 2 | $x_1^2$ | $x_2^2$ | $\cdots$ | $x_i^2$ | $\cdots$ | $x_n^2$ | $y^2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| j | $x_1^j$ | $x_2^j$ | $\cdots$ | $x_i^j$ | $\cdots$ | $x_n^j$ | $y^j$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| m | $x_1^m$ | $x_2^m$ | $\cdots$ | $x_i^m$ | $\cdots$ | $x_n^m$ | $y^m$ |

First, we introduce multiple linear regression.

**Definition 8 (Multiple Linear Regression Model).** *Let $a_0, a_1, \ldots, a_n \in \mathbb{R}$. Multiple linear regression model is defined by*

$$y = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + a_0.$$

*$a_0$ is the intercept point.*

Multiple linear regression model is the most common type model of regression analyses. But it does not have high describing ability, so that several models are proposed. Especially, models used by monotone measure are able to describe the interaction between explanatory variables. Let $(x_1, x_2, \ldots, x_n, y)$ be a data, $v$ be a monotone measure on $2^N$ and $m^v$ be a Möbius transform of $v$. Put a function $f := (x_1, x_2, \ldots, x_n)$ defined on $N$.

**Definition 9 (Choquet Integral Model[15]).** *Let $a_0, a_1, \ldots, a_n, a_{1,2}, \ldots, a_{1,2,\ldots,n} \in \mathbb{R}$. Choquet integral model is defined by*

$$y = (C)\int f dv = \sum_{A \subset N} \left\{ \sum_{B \supset A} \left( (-1)^{|B \setminus A|} \bigwedge_{i \in B} f(i) \right) \right\} v(A)$$

$$= a_1 \sum_{I \supset \{1\}} \left( (-1)^{|I \setminus \{1\}|} \bigwedge_{i \in I} x_i \right) + \cdots + a_n \sum_{I \supset \{n\}} \left( (-1)^{|I \setminus \{n\}|} \bigwedge_{i \in I} x_i \right)$$

$$= a_{1,2} \sum_{I \supset \{1,2\}} \left( (-1)^{|I \setminus \{1,2\}|} \bigwedge_{i \in I} x_i \right) + \cdots + a_{n-1,n} \sum_{I \supset \{n-1,n\}} \left( (-1)^{|I \setminus \{n-1,n\}|} \bigwedge_{i \in I} x_i \right)$$

$$\vdots$$

$$+ a_{1,2,\ldots,n}(x_1 \wedge x_2 \wedge \cdots \wedge x_n) + a_0,$$

*and Möbius type Choquet IE integral model is as follows:*

$$y = (C) \int f dv = \sum_{A \subset N} \left( \bigwedge_{i \in A} f(i) \right) m^v(A)$$

$$= a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

$$+ a_{1,2}(x_1 \wedge x_2) + a_{1,3}(x_1 \wedge x_3) + \cdots + a_{n-1,n}(x_{n-1} \wedge x_n)$$

$$+ a_{1,2,3}(x_1 \wedge x_2 \wedge x_3) + a_{1,2,4}(x_1 \wedge x_2 \wedge x_4) + \cdots + a_{n-2,n-1,n}(x_{n-2} \wedge x_{n-1} \wedge x_n)$$

$$\vdots$$

$$+ a_{1,2,\ldots,n}(x_1 \wedge x_2 \wedge \cdots \wedge x_n) + a_0.$$

The definitions of multiplication IE integral model are similar to Choquet IE integral model. It is IE integral model adopted multiplication as $n$-are operation. This integral corresponds with the multi-linear extension defined by G. Owen[16].

**Definition 10 (Multiplication IE Integral Model).** *Let* $a_0, a_1, \ldots, a_n, a_{1,2}, \ldots, a_{1,2,\ldots,n} \in \mathbb{R}$. *Multiplication IE integral model is defined by*

$$y = \Pi \int f dv = \sum_{A \subset N} \left\{ \sum_{B \supset A} \left( (-1)^{|B \setminus A|} \prod_{i \in B} f(i) \right) \right\} v(A)$$

$$= a_1 \sum_{I \supset \{1\}} \left( (-1)^{|I \setminus \{1\}|} \prod_{i \in I} x_i \right) + \cdots + a_n \sum_{I \supset \{n\}} \left( (-1)^{|I \setminus \{n\}|} \prod_{i \in I} x_i \right)$$

$$= a_{1,2} \sum_{I \supset \{1,2\}} \left( (-1)^{|I \setminus \{1,2\}|} \prod_{i \in I} x_i \right) + \cdots + a_{n-1,n} \sum_{I \supset \{n-1,n\}} \left( (-1)^{|I \setminus \{n-1,n\}|} \prod_{i \in I} x_i \right)$$

$$\vdots$$

$$+ a_{1,2,\ldots,n}(x_1 x_2 \cdots x_n) + a_0,$$

*and Möbius type Multiplication IE integral model is as follows:*

$$y = \Pi \int f dv = \sum_{A \subset N} \left( \prod_{i \in A} f(i) \right) m^v(A)$$

$$= a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

$$+ a_{1,2}(x_1 \cdot x_2) + a_{1,3}(x_1 \cdot x_3) + \cdots + a_{n-1,n}(x_{n-1} \cdot x_n)$$

$$+ a_{1,2,3}(x_1 \cdot x_2 \cdot x_3) + a_{1,2,4}(x_1 \cdot x_2 \cdot x_4) + \cdots + a_{n-2,n-1,n}(x_{n-2} \cdot x_{n-1} \cdot x_n)$$

$$\vdots$$

$$+ a_{1,2,\ldots,n}(x_1 \cdot x_2 \cdots x_n) + a_0.$$

The definitions of Dombi's IE integral model are also similar to other IE integral models, but Dombi's $T$-norm is defined on $[0,1]^2$, so that we have to normalize explanatory values within $[0,1]$.

**Definition 11 (Dombi's $T$-norm IE Integral Model).** *Let* $a_0, a_1, \ldots, a_n, a_{1,2}, \ldots, a_{1,2,\ldots,n} \in \mathbb{R}$. *Let* $f' = (x'_1, x'_2, \ldots, x'_n)$ *be a* $[0,1]$ *normalized* $f$, *in other words,*

$$f'(i) = \frac{f(i) - \bigwedge f(i)}{\bigvee f(i) - \bigwedge f(i)},$$

*where there are* $m$ *data set and* $\bigvee f(i)$ *and* $\bigwedge f(i)$ *are the maximum and the minimum values of* $m$ $f(i)$. *Then Dombi's integral model is defined by Dombi's IE integral model is defined by*

$$\begin{aligned}
y =& (D)\int f'dv = \sum_{A\subset N}\left\{\sum_{B\supset A}\left((-1)^{|B\setminus A|}\bigodot_{i\in B} f'(i)\right)\right\} v(A)\\
=& a_1\sum_{I\supset\{1\}}\left((-1)^{|I\setminus\{1\}|}\bigodot_{i\in I} x'_i\right) + \cdots + a_n\sum_{I\supset\{n\}}\left((-1)^{|I\setminus\{n\}|}\bigodot_{i\in I} x'_i\right)\\
=& a_{1,2}\sum_{I\supset\{1,2\}}\left((-1)^{|I\setminus\{1,2\}|}\bigodot_{i\in I} x'_i\right) + \cdots + a_{n-1,n}\sum_{I\supset\{n-1,n\}}\left((-1)^{|I\setminus\{n-1,n\}|}\bigodot_{i\in I} x'_i\right)\\
&\vdots\\
&+a_{1,2,\ldots,n}(x'_1\odot x'_2\odot\cdots\odot x'_n) + a_0,
\end{aligned}$$

*where* $\odot$ *is Dombi's* $T$*-norm, and Möbius type Dombi's IE integral model is as follows:*

$$\begin{aligned}
y =& (D)\int f'dv = \sum_{A\subset N}\left(\bigodot_{i\in A} f'(i)\right) m^v(A)\\
=& a_1x'_1 + a_2x'_2 + \cdots + a_nx'_n\\
&+a_{1,2}(x'_1\odot x'_2) + a_{1,3}(x'_1\odot x'_3) + \cdots + a_{n-1,n}(x'_{n-1}\odot x'_n)\\
&+a_{1,2,3}(x'_1\odot x'_2\odot x'_3)+a_{1,2,4}(x'_1\odot x'_2\odot x'_4) + \cdots + a_{n-2,n-1,n}(x'_{n-2}\odot x'_{n-1}\odot x'_n)\\
&\vdots\\
&+a_{1,2,\ldots,n}(x'_1\odot x'_2\odot\cdots\odot x'_n) + a_0.
\end{aligned}$$

Concerning to Möbius type models, assuming $v$ $k$-additive measure, we can reduce the number of terms from $2^n$ to $\sum_{\ell=1}^{k} {}_nC_\ell$.

## 3 Rescaling of Data – Our Proposal Methods

In this section, we discuss preprocessing of input data, which are explanatory variables in our case. Explanatory variables have deviation in the distribution depending on each special quality. We give an example.

*Example 1. On subjective evaluation of video quality, a model has been recommended by Telecommunication Standardization Sector of the International Telecommunication Union(ITU-U) as J.247 Annex A[17]. It's explanatory values are*

$$\begin{aligned} x_1 &:= \text{overall noize},\\ x_2 &:= \text{degradation caused by block distortion},\\ x_3 &:= \text{degradation associated with blurring},\\ x_4 &:= \text{local spatial degradation},\\ x_5 &:= \text{freeze degradation}, \end{aligned}$$

*which all are objective values and human subjective evaluation of video quality $y$ is predicted by*

$$\begin{aligned} y &= a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5\log_{10}(x_5)\\ &\qquad +a_6(a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4)\log_{10}(x_5) + a_0\\ &= a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5' \\ &\qquad +(a_1a_6)x_1x_5' + (a_2a_6)x_2x_5' + (a_3a_6)x_3x_5' + (a_4a_6)x_4x_5' + a_0, \end{aligned} \tag{1}$$

*where $x_5' := \log_{10}(x_5)$.*

We can regard the model as the multiple IE integral model with 2-additive measure considering that coefficients of $x_2x_3, x_2x_4, \ldots, x_3x_4$ are vanished. The model specialize in subjective video quality estimation and is based on experience. Explanatory variable $x_5$ is rescaled by function $Res(x) = \log_{10} x$. Values of $x_5$ largely concentrate to a small interval so that rescaling them by the increasing function $\varphi(x) = \log_{10} x$ is reasonable for the sake of dispersing them. We propose two methods for determining increasing rescaling function using training data of explanatory values.

**Definition 12 (Our Proposed Method 1, Rearrangement Method).** *Let $x^1, x^2, \ldots, x^m$ be training data of an explanatory value.*

1. *Rearrange $x^1, x^2, \ldots, x^m$ in ascending order, that is, let $x^1 \le x^2 \le \cdots \le x^m$.*
2. *Let $\delta := (x^m - x^1)/(m-1)$.*
3. *Let $m+2$ points $X^0 := (x^1 - \delta, 0), X^1 := (x^1, 1), X^2 := (x^2, 2), \cdots, X^m := (x^m, m), X^{m+1} := (x^1 + \delta, m+1)$. If $x^j = \cdots = x^{j+\ell}$, then let $X^j = \cdots = X^{j+k} := (x_j, j + k/2)$.*
4. *Define $Res(x)$ on $\mathbb{R}$ by $X^0, \cdots, X^m$ and their linear interpolation.*

**Definition 13 (Our Proposed Method 2, Cumulative Frequency Distribution Method).** *Let $x^1, x^2, \ldots, x^m$ be training data of an explanatory value.*

1. *Rearrange $x^1, x^2, \ldots, x^m$ in ascending order, that is, let $x^1 \le x^2 \le \cdots \le x^m$.*
2. *Make a graph of cumulative frequency distribution of data.*
3. *Let 2 points $X^0 := (2x^1 - x^m, 0), X^{m+1} := (2x^n - x^1, m+1)$.*
4. *Define $Res(x)$ on $\mathbb{R}$ by the graph of cumulative frequency distribution, $X^0$, $X^m$ and their linear interpolation.*

## 4 Experimentation

We analyze real data sets of human subjective evaluations on video quality in Example 1 to verify the validity of our proposal methods. We prepare 1139 data sets of human subjective evaluations on video quality. Each data set $(y, x_1, x_2, x_3, x_4, x_5)$ is consist of one response variable which is subjective evaluations on video quality and five explanatory values, that is, $n = 5$ . Using 500 data sets of 1139 data sets, we determine each values of the parameters, then we estimate values of response variable using 500 data sets of other 639 data sets.

### 4.1 Regression Analyses Models

We adopt the following four models.

**1. MLR Model** multiple linear regression model:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_0.$$

**2. 2-MCIEI Model** 2-additive Möbius type Choquet IE integral model:

$$\begin{aligned} y = {} & a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 \\ & + a_{1,2}(x_1 \wedge x_2) + a_{1,3}(x_1 \wedge x_3) + a_{1,4}(x_1 \wedge x_4) + a_{1,5}(x_1 \wedge x_5) \\ & + a_{2,3}(x_2 \wedge x_3) + a_{2,4}(x_2 \wedge x_4) + a_{2,5}(x_2 \wedge x_5) + a_{3,4}(x_3 \wedge x_4) \\ & + a_{3,5}(x_3 \wedge x_5) + a_{4,5}(x_4 \wedge x_5) + a_0. \end{aligned} \tag{2}$$

**3. 2-MMIEI Model** 2-additive Möbius type multiplication IE integral model which is the model replaced $\wedge$ with multiplicative operator in (2).

**4. 2-MDIEI Model** 2-additive Möbius type Dombi's IE integral model which is the model replaced $\wedge$ with Dombi's $T$-norm, $\odot$ in (2). We set the parameter $\lambda = 2.5$. Explanatory values, $x_1, \ldots, x_5$ have to be noralized in $[0, 1]$.

### 4.2 Rescaling Methods

We attempt non-rescaling and two rescaling methods, rearrangement method (Rearrange method) and cumulative frequency distribution method(CFD method) for each models. We also Therefore there are twelve pattern of combinations of models and rescaling methods. Besides we also adopt the model ITU-U recommended (1).

### 4.3 Results

Using 500 data sets, we determine each values of the parameters which minimize mean square error, then we estimate values of response variable using other 500 data sets. We remark that obtained values as the parameters of 2-MCIEI model,

2-MMIEI model, 2-MMIEI model and 2-MDIEI model are not generally non-additive measure. They don't satisfy non-negativity or monotonicity in general. We compare these nine methods with mean square error(MSE) of each estimated values $\hat{y}^j$ and corresponded response variables of data $y^j$:

$$MSE = \frac{1}{500} \sum_{j=1}^{500} \frac{(y^j - \hat{y}^j)^2}{y^j}.$$

**Table 2.** Mean square error(MSE)

| | MLR | 2-MCIEI | 2-MMIEI | 2-MDIEI | ITU-U |
|---|---|---|---|---|---|
| num. of parameters | 6 | 16 | 16 | 16 | 6 |
| non-rescaling | 0.3917 | 0.3037 | 0.3176 | 0.2798 | |
| Method 1 Rearrange | 0.3065 | 0.2614 | 0.2798 | 0.2885 | 0.2938 |
| Method 2 CHD | 0.2976 | 0.2422 | 0.2577 | 0.2493 | |

Our result is shown in table 2 which indicates the validity of rescaling by our both proposed methods comparing with non-rescaling and also ITU-U model. ITU-U model is empirically obtained. And whereas our rescaling methods enable us to construct a precise model in algorithmically way.

## 5 Conclusions

We have examined the validity of the proposed methods applying to the estimation for subjective evaluation with large reliable data. We may go on from the results of the examination to the conclusion that our proposed rescaling methods are valid for any regression analyses models. Especially, our rescaling methods is considered that an excellent effect is obtained in the Choquet integral model, because in the case of the Choquet integral model,

1. it can be difficult to retain independency among values. For example, in an extreme case that all data satisfies $x_1 < x_2 < \cdots < x_n$, then the values $f(1), f(1) \wedge f(2), \ldots, f(1) \wedge \cdots \wedge f(n)$ are completely same. In the case that there are some data depending on other data, we cannot estimate the parameters at all. And
2. minimum operation of explanatory values means their comparison, so that we have to make their scales same range. Using method 2, it is automatically realized.

The future plans are

1. studying properties of operators using in IE integral model, and find well-suited operators to our two rescaling methods. and
2. applying our methods to various real data and also investigating other conventional models of various subjective evaluation problems which can be interpreted as rescaling to verify validity of our rescaling methods.

## References


1. Grabisch, M.: Fuzzy pattern recognition by fuzzy integrals and fuzzy rules. In: Pal, S.R., Pal, A. (eds.) Om: Pattern Recognition: From Classical to Modern Approaches World Scientific, pp. 257–280 (2002)
2. Sugeno, M.: Fuzzy measures and fuzzy integrals–a survey. In: Gupta, M.M., Saridis, G.N., Gaines, B.R. (eds.) Fuzzy Automata and Decision Processes, pp. 89–102. North Holland, Amsterdam (1977)
3. Choquet, G.: Theory of capacities. Ann. Inst. Fourier 5, 131–295 (1953)
4. Grabisch, M., Labreuche, C.: A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. A Quarterly Journal of Operations Research 6, 1–44 (2008)
5. Grabisch, M.: k-order additive discrete fuzzy measures and their representation. Fuzzy Sets and Systems 92, 167–189 (1991)
6. Mori, D., Honda, A., Uchida, M., Okazaki, Y.: Quality evaluations of network services using a non-additive set function. In: Proceedings of Modeling Decisions for Artificial Intelligence (CD-ROM) (2008)
7. Honda, A., Okamoto, J.: Inclusion-exclusion integral and its application to subjective video quality estimation. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) IPMU 2010. CCIS, vol. 80, pp. 480–489. Springer, Heidelberg (2010)
8. Mesiar, R.: Choquet-like integral. J. Math Anal. Appl. 194, 477–488 (1995)
9. Benvenuti, P., Vivona, D.: General theory of the fuzzy integral. Mathware and Soft Computing 3, 199–209 (1996)
10. Klement, E.P., Kolesarova, A., Mesiar, R.: Lipschitz continuity of discrete universal integrals based on copuras. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 18(1), 39–52 (2010)
11. Schweizer, B., Sklar, A.: Statistical metric spaces. Pacific. J. Math. 10, 313–334 (1960)
12. Klement, E.P.: Some mathematical aspects of fuzzy sets: Triangular norms, fuzzy logics, and generalized measures. Fuzzy Sets Syst. 90, 133–140 (1997)
13. Mesiar, R., Pap, E.: Different interpretations of triangular norms and related operations. Fuzzy Sets and Systems 96, 183–189 (1998)
14. Dombi, J.: A general class of fuzzy operators, the DeMorgan class of fuzzy operators and fuzziness measures induced by fuzzy operators. Fuzzy Sets Syst. 8, 149–163 (1982)
15. Torra, V., Narukawa, Y.: Modeling decisions: information fusion and aggregation operators. Springer (2007)
16. Owen, G.: Multilinear extensions of games. Management Sci. 18(5), 64–79 (1972)
17. ITU-T Recommendation J.247: Objective perceptual multimedia video quality measurement in the presence of a full reference (2008)

# Construction of a Bi-capacity and Its Utility Functions without any Commensurability Assumption in Multi-criteria Decision Making

Christophe Labreuche

Thales Research & Technology
1 avenue Augustin Fresnel
91767 Palaiseau Cedex - France

**Abstract.** We consider a multi-criteria evaluation function $U$ defined over a Cartesian product of attributes. We assume that $U$ is written as the combination of an aggregation function and one value function over each attribute. The aggregation function is assumed to be a Choquet integral w.r.t. an unknown bi-capacity. The problem we wish to address in this paper is the following one: if $U$ is known, can we construct both the value functions and the bi-capacity? The approaches that have been developed so far in the literature to answer this question in an analytical way assume some commensurability hypothesis. We propose in this paper a method to construct the value functions and the capacity without any commensurability assumption. Moreover, we show that the construction of the value functions is unique up to an affine transformation.

**Keywords:** Choquet integral, bi-capacity, commensurability, bipolarity.

## 1 Introduction

One of the most widely used model in decision theory is the weighted sum. The overall value is then written as a weighted sum of value functions on each attribute. This model has been extensively studied in the context of decision under uncertainty [1,2] and multi-criteria decision making [3]. The main drawback of this model is that it assumes the independence among the criteria. The Choquet integral which subsumes the weighted sum, has been used in decision theory to overcome the limitation of the weighted sum [4]. The Choquet integral is a versatile aggregation function that has been applied to Multi-Criteria Decision Analysis [5] and Data fusion. Its main interest lies in its ability to represent interaction among the attributes. The overall value for this model is written as the Choquet integral of value functions on each attribute. The model of the Choquet integral has been extended to represent bipolar decision behaviour [6]. Conditional relative importance is an example of such decision strategy. A bipolar scale is characterized by a neutral element that is neither good nor bad. A bipolar Choquet integral is described by a bi-capacity.

Unlike the weighted sum, the use of the Choquet integral implies that the scales representing the attributes are commensurate. There are very few works

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 294–303, 2014.

in which both the Choquet integral and the value functions are constructed in an analytical way without making any a priori assumption about commensurability. One can mention reference [7] which considers the case of a Choquet integral. The aim of this paper is to extend this approach to bipolar Choquet integrals.

In [8], the neutral element of a bipolar scale is determined under the knowledge that the decision maker satisfies a special decision strategy called conditional relative importance between criteria. In this work, by analyzing the non linear decision strategies represented in the overall utility function, it is possible to identify the neutral value in the value functions. This idea will be used and extended in this paper.

The main idea of this paper is now summarized. Commensurability between the criteria is required for the bipolar Choquet integral since this aggregation function is based on a ranking of the values of the criteria. Yet, the bipolar Choquet integral is a piecewise weighted sum function, in which the weights of the criteria are the same whenever the criteria keep the same sign and are ranked in the same order. Considering two criteria $i$ and $k$, the weight of criterion $i$ depends on the relative utilities on criteria $i$ and $k$. This means that, if the value $x_k$ of criterion $k$ varies and the other criteria are fixed, then one may observe that the weight of criterion $i$ suddenly changes when the utility on criterion $k$ is equal to the utility on criterion $i$ or to the opposite of the utility on criterion $i$. Hence, from an element $x_i$ on attribute $i$, one can identify two elements $x_k^1$ and $x_k^2$. The neutral level over an attribute $X_i$ can be uniquely identified as the single value for which $x_k^1 = x_k^2$. This construction does not work if the weight of criterion $i$ does not depend on criterion $k$. If this holds for any value on the other criteria, one can show that the criteria $i$ and $k$ are independent. Applying this construction to any pair $i, k$ of criteria, one obtains a partition of the set of criteria. In each set, the criteria interact one with another, and it is thus possible to construct vectors of values on the attributes that are commensurate. There is complete independence between the criteria of any two sets in this partition. Hence one cannot ensure commensurability between two sets in the partition. But this is not a problem since the bipolar Choquet integral is additive between groups of criteria that are independent. When the neutral elements are obtained and commensurate reference elements are identified, then the utility functions and the bi-capacity can be derived by extending the MACBETH methodology [9].

The layout of the paper is as follows. Our major difficulty in the construction of the value function is to construct two reference values on each attribute that are commensurate. This is done in Section 4, based on the construction that we have just presented. When two sequences of commensurate values are obtained, we are able to construct the value functions and the bi-capacity by generalizing the approach of [10] to the case when the commensurability is attained only within each set in a partition of the criteria (see Section 5).

## 2 Notation and the Choquet Integral on a Bipolar Scale

Consider a problem of selecting one option among several, where each option is described by several attributes. $N = \{1, \ldots, n\}$ is the set of attributes and

the set of possible values of attribute $i \in N$ is denoted by $X_i$. Options are thus elements of the product set $X := X_1 \times \cdots \times X_n$. We are looking at a numerical representation $U$ of the option in the following form:

$$U(x) = F(u_1(x_1), \ldots, u_n(x_n)) \tag{1}$$

where $F : \mathbb{R}^N \to \mathbb{R}$ is the aggregation function and $u_i : X_i \to \mathbb{R}$ are the utility functions.

### 2.1 Choquet Integral

A *capacity* (also called *fuzzy measure*) on a set $N$ of criteria is a set function $\nu : 2^N \to [0,1]$ such that $\nu(\emptyset) = 0$, $\nu(N) = 1$ and $\nu(A) \leq \nu(B)$ for all $A \subseteq B \subseteq N$.

The Choquet integral of $x = (x_1, \ldots, x_n) \in [0,1]^n$ defined w.r.t. a game $\nu$ has the following expression [11] :

$$C_\nu(a_1, \ldots, a_n) = \sum_{i=1}^{n} \left(a_{\tau(i)} - a_{\tau(i-1)}\right) \nu\left(\{\tau(i), \cdots, \tau(n)\}\right), \tag{2}$$

where $a_{\tau(0)} := 0$ and $\tau$ is a partition on $N$ such that $a_{\tau(1)} \leq a_{\tau(2)} \leq \cdots \leq a_{\tau(n)}$. The Choquet integral has been proved to be able to model both the importance of criteria and the interaction between criteria.

### 2.2 Bi-capacity

A bipolar scale is characterized by a neutral element, demarcating attractive and repulsive elements. Th neutral level on attribute $i$ is noted $\mathbb{O}_i$. It is characterized by $u_i(_zero_i) = 0$. Let

$$\mathcal{Q}(N) = \{(S,T) \in \mathcal{P}(N) \times \mathcal{P}(N) \mid S \cap T = \emptyset\}$$

A bi-capacity is a function $\mu : \mathcal{Q}(N) \to \mathbb{R}$ satisfying [12]: (i) $S \subseteq S' \Rightarrow \mu(S,T) \leq \mu(S',T)$; (ii) $T \subset T' \Rightarrow \mu(S,T) \geq \mu(S,T')$; (iii) $\mu(\emptyset,\emptyset) = 0$, $\mu(N,\emptyset) = 1$, $\mu(\emptyset,N) = -1$.

The first two properties depict increasingness. $\mu(S,T)$ is interpreted as the overall assessment of the ternary act $(1_S, -1_T, 0_{-S\cup T})$ taking value 1 on attributes in $S$, value $-1$ on attributes in $T$ and value 0 on the remaining attributes.

The Choquet integral w.r.t. a bi-capacity has been proposed in [12]. Let $a \in \mathbb{R}^N$, $N^+ = \{i \in N, a_i \geq 0\}$ and $N^- = N \setminus N$. Define the capacity $\nu$ by

$$\forall S \subseteq N \ , \ \nu(S) := \mu\left(S \cap N^+, S \cap N^-\right) \ .$$

Then the Choquet integral w.r.t. $\mu$ is defined by:

$$BC_\mu(a) := C_\nu\left(a_{N^+}, -a_{N^-}\right) \ . \tag{3}$$

## 3 Statement of the Problem

In this paper, we assume that we know the overall value $U(x)$ for all $x \in X$. It is supposed to take the form for all $x \in X$,

$$U(x) = BC_\mu(u_1(x_1), \ldots, u_n(x_n)). \tag{4}$$

The aim is to construct $\mu$ and $u_1, \ldots, u_n$ without knowing $\mathbb{O}_{1,l}\, dotrs, \mathbb{O}_n$.

For the sake of simplicity, we assume that

$$X = \mathbb{R}^n,$$

and that the value functions are strictly increasing. Furthermore, we assume that $U$ is continuous and that for all $i \in N$, there exists $x_{-i} \in X_{-i}$ such that

$$\lim_{x_i \to +\infty} U(x) = +\infty \text{ and } \lim_{x_i \to -\infty} U(x) = -\infty. \tag{5}$$

## 4 Properties on Profiles of Commensurate Values

Consider $i, k \in N$ with $i \neq k$. We assume here that we know $\mathbb{O}_i$. We wish to find some properties allowing us to find $x_i, x_k$ that are *commensurate* (i.e. such that $u_i(x_i) = u_k(x_k)$) or *skew-commensurate* (i.e. such that $u_i(x_i) = -u_k(x_k)$). Set $\delta_i^{\pm}$ by

$$\delta_i^+ \mu(C^+, C^-) = \mu(C^+ \cup \{i\}, C^-) - \mu(C^+, C^-)$$
$$\delta_i^- \mu(C^+, C^-) = \mu(C^+, C^- \cup \{i\}) - \mu(C^+, C^-).$$

If $U$ follows expression (4), then we have

$$\frac{\partial U}{\partial x_i}(x) = \begin{cases} \varepsilon\, \delta_i^\varepsilon \mu(C^+ \cup \{k\}, C^-)\, u_i'(x_i) & \text{if } u_k(x_k) > u_i(x_i) \\ \varepsilon\, \delta_i^\varepsilon \mu(C^+, C^-)\, u_i'(x_i) & \text{if } u_i(x_i) > u_k(x_k) > -u_i(x_i) \\ \varepsilon\, \delta_i^\varepsilon \mu(C^+, C^- \cup \{k\})\, u_i'(x_i) & \text{if } -u_i(x_i) > u_k(x_k) \end{cases}$$

where $\varepsilon = +$ if $u_i(x_i) > 0$, and $\varepsilon = -$ otherwise, and $C^+, C^- \in \mathcal{Q}(N \setminus \{i, k\})$.

We have the following result.

**Theorem 1.** *Let $x_i \in X_i$. Then one of the following three cases holds:*

*(i) for all $x_i \in X_i \setminus \{\mathbb{O}_i\}$, there exists $x_{-i,k} \in X_{-i,k}$ and $x_k^-, x_k^+ \in X_k$ such that the function $x_k \mapsto \frac{\partial U}{\partial x_i}(x)$ takes 3 values and is constant in the three intervals $(-\infty, x_k^-)$, $(x_k^-, x_k^+)$ and $(x_k^+, \infty)$. Values $x_k^+$ and $x_k^-$ satisfy*

$$\begin{cases} u_k(x_k^+) = u_i(x_i) \text{ and } u_k(x_k^-) = -u_i(x_i) & \text{if } u_i(x_i) > 0 \\ u_k(x_k^+) = -u_i(x_i) \text{ and } u_k(x_k^-) = u_i(x_i) & \text{if } u_i(x_i) < 0 \end{cases} \tag{6}$$

*For $x_i = \mathbb{O}_i$, there exists $x_{-i,k} \in X_{-i,k}$ and $x_k^\star \in X_k$ such that the function $x_k \mapsto \frac{\partial U}{\partial x_i}(x)$ takes 2 values (and there does not exist $x_{-i,k} \in X_{-i,k}$ such that the function $x_k \mapsto \frac{\partial U}{\partial x_i}(x)$ takes 3 values) and is constant in the two intervals $(-\infty, x_k^\star)$ and $(x_k^\star, \infty)$. Then $u_k(x_k^\star) = 0$.*

*(ii) for all $x_i \in X_i$, there exists $x_{-i,k} \in X_{-i,k}$ and $x_k^\star \in X_k$ such that the function $x_k \mapsto \frac{\partial U}{\partial x_i}(x)$ takes 2 values (and there does not exist $x_{-i,k} \in X_{-i,k}$ such that the function $x_k \mapsto \frac{\partial U}{\partial x_i}(x)$ takes 3 values) and is constant in the two intervals $(-\infty, x_k^\star)$ and $(x_k^\star, \infty)$. Then*

$$u_k(x_k^\star) \in \{-u_i(x_i), u_i(x_i)\}. \tag{7}$$

*(iii) for all $x_i \in X_i$ and all $x_{-i,k} \in X_{-i,k}$, the function $x_k \mapsto \frac{\partial U}{\partial x_i}(x)$ is constant. Then $\forall \epsilon \in \{+,-\}$ and $\forall (C^+, C^-) \in \mathcal{Q}(N \setminus \{i,k\})$*

$$\delta_i^\epsilon \mu(C^+ \cup \{k\}, C^-) = \delta_i^\epsilon \mu(C^+, C^-) = \delta_i^\epsilon \mu(C^+, C^- \cup \{k\}). \tag{8}$$

The situation proved in Theorem 1 is summarized in Figure 1.

If we are in the situation (i) in Theorem 1, the two elements $x_k^-$ and $x_k^+$ that are constructed from $x_i$ are denoted by $\mathrm{Comm}^-_{i\to k}(x_i)$ and $\mathrm{Comm}^+_{i\to k}(x_i)$, where Comm stands for commensurability. We are indeed able to find an element on attribute $k$ that is commensurate with the element $x_i$ on attribute $i$. Moreover it is also apparent that $x_k^\star = \mathbb{O}_k$.

Situation (ii) corresponds to a degenerate bipolar case. This occurs for instance if we consider the Choquet integral w.r.t. a standard capacity.

Situation (iii) in Theorem 1 implies that criteria $i$ and $k$ are independent (see (8)). This is an important consequence of this theorem.

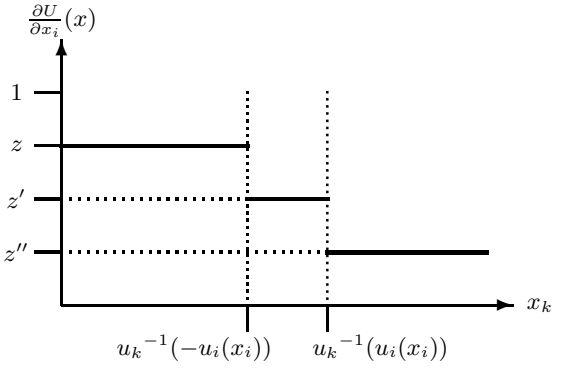



**Fig. 1.** Function $x_k \mapsto \frac{\partial U}{\partial x_i}(x)$ with $z := \varepsilon\, \delta_i^\varepsilon \mu(C^+, C^- \cup \{k\})\, u_i'(x_i)$, $z' := \varepsilon\, \delta_i^\varepsilon \mu(C^+, C^-)\, u_i'(x_i)$ and $z'' := \varepsilon\, \delta_i^\varepsilon \mu(C^+ \cup \{k\}, C^-)\, u_i'(x_i)$

From now on, we will assume that for every pair of criteria, only the two situations (i) and (iii) can occur. Case (ii) is left for future research.

## 5 Construction of the Bi-capacity and the Value Functions

We propose a step by step construction of the bi-capacity and the utility functions, assuming that $U$ is known. Following [10], it is possible to construct the

value functions on a bounded unipolar scale through a Choquet integral with an unknown capacity as soon as one can identify on each attribute two reference levels $\mathbb{O}$ and $\mathbb{G}$ ($\mathbb{G}$ stands for *Good*) on each attribute. For unbounded bipolar scales, one needs three reference levels $\mathbb{B}$ ($\mathbb{B}$ stands for *Bad*), $\mathbb{O}$ (the neutral and $\mathbb{G}$ to construct a value function [6], with the property that $u_i(\mathbb{G}_i) = 1 = -u_i(\mathbb{B}_i)$.

### 5.1 Determination of the Partition of Criteria

First of all, according to Theorem 1, two criteria $i$ and $k$ can either be make commensurate (case (i)) or not (case (iii)). Hence the set $N$ of criteria can be partitioned into disjoint subsets such that one can construct commensurate profiles in each subset. We introduce the Algorithm 1 to construct a partition $\mathcal{T}$ of $N$.

```
Function getPartition() :
    T ← ∅ and R ← N;
    While (R ≠ ∅) do
        Select j ∈ R; S ← {j} and R ← R \ {j};
        While (∃i ∈ S, k ∈ R such that situation (i) in Theorem 1 holds) do
            S ← S ∪ {k}; R ← R \ {k};
        done
        T ← {T, S};
    done
    return T;
End
```

**Algorithm 1.** Algorithm for the determination of the partition $\mathcal{T}$

### 5.2 Determination of the Neutral Elements

By Theorem 1, if $i, k$ are in situation (i), then the neutral element $\mathbb{O}_i$ can be easily characterized: it is the unique $x_i \in X_i$ such that $\text{Comm}^-_{i\to k}(x_i) = \text{Comm}^+_{i\to k}(x_i)$. One can easily design a dichotomy algorithm that returns an estimate of the neutral element. The precise description of this algorithm is omitted due to space limitation.

When $S = \{i\}$ with $S \in \mathcal{T}$, then the concept of a neutral element does not exist and we fix $\mathbb{O}_i$ to any element in $X_i$.

### 5.3 Construction of the Value Functions on the Positive Part of the Scales

We define $\mathbb{G}, \mathbb{B}$ as any output of **getCommensurateSequence**$(\mathcal{T}, \mathbb{O})$ (see Algorithm 2). In the algorithm, **getValue**$(j, \mathbb{O}_j)$ returns a random element of $X_j$ above $\mathbb{O}_j$. Theorem 2 below shows that the outputs of the algorithm have the required property for being commensurate profiles with one above the neutral element and the other one below the neural element.

**Function** getCommensurateSequences($\mathcal{T}, \mathbb{O}$) :
  **For all** $S \in \mathcal{T}$ **do**
    Select $i \in S$;
    $x_i \leftarrow$ **getValue**$(j, \mathbb{O}_j)$;
    **If** $(|S| = 1)$ **then**
      $y_i \leftarrow$ null;
    **else**
      $R \leftarrow \{i\}$;
      **While** $(\exists i \in R, k \in S \setminus R$ s.t. situation (i) in Theorem 1 holds) **do**
        $R \leftarrow R \cup \{k\}$;
        $x_k \leftarrow \mathrm{Comm}^{+}_{i \to k}(x_i)$;
        $y_k \leftarrow \mathrm{Comm}^{-}_{i \to k}(x_i)$;
        **If** $(|R| = 2)$ **then**
          $y_i \leftarrow \mathrm{Comm}^{-}_{k \to i}(x_k)$;
        **end If**
      **done**
    **end If**
  **end For**
  **return** $(x, y)$;
**End**

**Algorithm 2.** Determination of the commensurate reference profiles $\mathbb{G}, \mathbb{B}$

**Theorem 2.** *Let $\mathcal{T}$ be the outcome of the function "getPartition" (see Algorithm 1), and $(x, y)$ be the outcome of the function "getCommensurateSequence" (see Algorithm 2). Then $\mathcal{T}$ forms a partition of $N$. For all $S \in \mathcal{T}$, we have if $|S| > 1$*

$$\begin{cases} \forall i, k \in S \quad u_i(x_i) = u_k(x_k) > 0, \\ \forall i, k \in S \quad u_i(y_i) = u_k(y_k) < 0. \end{cases}$$

*Moreover if $\mu$ is a bi-capacity representing $U$, then*

$$\forall (C^+, C^-) \in \mathcal{Q}(N) \quad \mu(C^+, C^-) = \sum_{S \in \mathcal{T}} \mu(C^+ \cap S, C^- \cap S).$$

Consider first $i \in S \in \mathcal{T}$ with $|S| > 1$. Then the utility function on the positive part of the scale is define as follows: for all $x_i \in X_i$ with $x_i \geq \mathbb{O}_i$

$$v_i^+(x_i) = \frac{U(x_i, \mathbb{O}_{-i}) - U(\mathbb{O}_N)}{U(\mathbb{G}_i, \mathbb{O}_{-i}) - U(\mathbb{O}_N)}. \tag{9}$$

With this definition, we have the following normalization

$$v_i^+(\mathbb{G}_i) = 1 \quad \text{and} \quad v_i^+(\mathbb{O}_i) = 0.$$

The utility function on the negative part of the scale is defined as follows: for all $x_i \in X_i$ with $x_i \leq \mathbb{O}_i$

$$v_i^-(x_i) = \frac{U(x_i, \mathbb{O}_{-i}) - U(\mathbb{O}_N)}{U(\mathbb{O}_N) - U(\mathbb{B}_i, \mathbb{O}_{-i})}. \tag{10}$$

With this definition, we have the following normalization

$$v_i^-(\mathbb{B}_i) = -1 \quad \text{and} \quad v_i^-(\mathbb{O}_i) = 0.$$

Finally we set $v_i(x_i) = v_i^+(x_i)$ if $x_i \geq 0$ and $v_i(x_i) = v_i^-(x_i)$ otherwise.

When $S = \{i\}$, formula (9) is used to define $v_i(x_i)$ for all $x_i \in X_i$.

### 5.4 Identification of the Bi-capacity

For each $S \in \mathcal{T}$, we define a bi-capacity $m_S$ on $S$, from $\mathbb{B}, \mathbb{O}, \mathbb{G}$ by

$$\forall (C^+, C^-) \in \mathcal{Q}(S) \quad m_S(C^+, C^-) = \frac{U(\mathbb{G}_{C^+}, \mathbb{B}_{C^-}, \mathbb{O}_{-C^+ \cup C^-}) - U(\mathbb{O}_N)}{U(\mathbb{G}_S, \mathbb{O}_{-S}) - U(\mathbb{O}_N)}. \tag{11}$$

We set $m = \{m_S\}_{S \in \mathcal{T}}$.

### 5.5 Main Result

**Theorem 3.** *Let us consider an overall value function $U$ that takes the form (4), where $u_1, \ldots, u_n, \mu$ (and also $\mathbb{O}_1, \ldots, \mathbb{O}_n$) are unknown.*

*Let $\mathcal{T}, \mathbb{O}, \mathbb{G}, \mathbb{B}$ as defined previously, $v_i$ defined by (9) and (10), and $m$ defined by (11). Then for all $x \in X$*

$$U(x) = U(\mathbb{O}_N) + \sum_{S \in \mathcal{T}} (U(\mathbb{G}_S, \mathbb{O}_{-S}) - U(\mathbb{O}_N))\, BC_{m_S}(v_S(x_S)),$$

*where $v_S(x_S) = (v_{i_1}(x_{i_1}), v_{i_2}(x_{i_2}), \ldots, v_{i_s}(x_{i_s}))$ if $S$ writes $S = \{i_1, i_2, \ldots, i_s\}$.*

*Let us consider two constructions $\langle \mathcal{T}, \mathbb{O}, \mathbb{G}, \mathbb{B}, v_1, \ldots, v_n, m \rangle$ and $\langle \mathcal{T}', \mathbb{O}', \mathbb{G}', \mathbb{B}', v_1', \ldots, v_n', m' \rangle$ from $U$. Then $\mathcal{T}' = \mathcal{T}$ and $m' = m$. Moreover, for all $S \in \mathcal{T}$ there exists $\alpha_S > 0$ and $\beta_S \in \mathbb{R}$ such that*

$$\text{If } |S| > 1, \qquad \forall i \in S \quad \forall x_i \in X_i \quad v_i'(x_i) = \alpha_S v_i(x_i), \tag{12}$$

$$\text{If } |S| = 1, \qquad \forall i \in S \quad \forall x_i \in X_i \quad v_i'(x_i) = \alpha_S v_i(x_i) + \beta_S \tag{13}$$

The previous result shows that the scales are ratio scales (given up to a dilation) only for subsets $S \in \mathcal{T}$ such that $|S| > 1$. All compatible scales are identical up to dilation. The renormalization are different in each set of the partition $\mathcal{T}$. Moreover, the capacities defined on each set of the partition are uniquely defined.

**Lemma 1.** *Let $\mathcal{T}$ be a partition of $N$, and $\mu$ be a bi-capacity on $N$. Assume that*

$$\forall (C^+, C^-) \in \mathcal{Q}(N) \quad \mu(C^+, C^-) = \sum_{S \in \mathcal{T}} \mu(C^+ \cap S, C^- \cap S).$$

*Then for all $a \in \mathbb{R}^N$*

$$BC_\mu(a) = \sum_{S \in \mathcal{T}} BC_\mu(a_S, 0_{-S}).$$

From the previous lemma, the Choquet integral is just the sum of the aggregation functions over each $S \in \mathcal{T}$. This is the form of the GAI network that generalizes the additive value model [13].

**Corollary 1.** *We consider a pair of commensurate profiles $\mathbb{O}, \mathbb{G}$ such that*

$$\sum_{S \in \mathcal{T}} (U(\mathbb{G}_S, \mathbb{O}_{-S}) - U(\mathbb{O}_N)) = 1. \tag{14}$$

*It is always possible to construct profiles satisfying this relation.*

*Under the condition of Theorem 3, let us define a bi-capacity $\widehat{\mu}$ over $N$ by*

$$\begin{aligned}
&\forall S \in \mathcal{T}\ \forall (C^+, C^-) \in \mathcal{Q}(S) \\
&\qquad \widehat{\mu}(C^+, C^-) = (U(\mathbb{G}_S, \mathbb{O}_{-S}) - U(\mathbb{O}_N))\, m_S(C^+, C^-), \\
&\forall (C^+, C^-) \in \mathcal{Q}(N) \quad \widehat{\mu}(C^+, C^-) = \sum_{S \in \mathcal{T}} \widehat{\mu}(C^+ \cap S, C^- \cap S),
\end{aligned}$$

*and we define $\widehat{v}_i$ by*

$$\widehat{v}_i(x_i) = v_i(x_i) + U(\mathbb{O}_N).$$

*Then $\widehat{\mu}$ is a capacity, and for all $x \in X$*

$$U(x) = BC_{\widehat{\mu}}(\widehat{v}_1(x_1), \ldots, \widehat{v}_n(x_n)).$$

*Let us consider two constructions $\langle \mathcal{T}, \mathbb{O}, \mathbb{G}, \mathbb{B}\widehat{v}_1, \ldots, \widehat{v}_n, \widehat{\mu} \rangle$ and $\langle \mathcal{T}, \mathbb{O}', \mathbb{G}', \mathbb{B}', \widehat{v}'_1, \ldots, \widehat{v}'_n, \widehat{\mu}' \rangle$ satisfying (14). Then for all $S \in \mathcal{T}$ there exists $\alpha_S > 0$ and $\beta_S \in \mathbb{R}$ such that*

$$\begin{aligned}
&\text{If } |S| > 1, \qquad \forall i \in S \quad \forall x_i \in X_i \quad \widehat{v}'_i(x_i) = \alpha_S \widehat{v}_i(x_i), \\
&\text{If } |S| = 1, \qquad \forall i \in S \quad \forall x_i \in X_i \quad \widehat{v}'_i(x_i) = \alpha_S \widehat{v}_i(x_i) + \beta_S, \\
&\widehat{\mu}'(S, \emptyset) = \frac{\widehat{\mu}(S, \emptyset)}{\alpha_S}.
\end{aligned}$$

Condition (14) is the normalization condition. It implies that the bi-cooperative game $\widehat{\mu}$ that is constructed, is a bi-capacity.

Concerning the second part of the previous corollary, it says that if the scale $\widehat{v}$ is a ratio scale when coalition $S \in \mathcal{T}$ contrains at least two criteria. Moreover, if $\widehat{v}$ is transformed, then the bi-capacity $\widehat{\mu}$ of $S$ shall be transformed by the inverse of this transformation. As a whole the overall utility thus remains the same.

## 6 Conclusion

One of the main difficulties in the application of the Choquet integral in multicriteria decision making is the necessity to make the value function commensurate. This paper proposes an approach to construct the Choquet integral without any prior commensurability assumption. The idea is that if the value of criterion k varies and the other criteria are fixed, then one may observe that the weight of criterion i suddenly changes when the value of criterion k is equal to that of criterion i. This enables us to construct sequences of commensurate values on the different attributes.

For future work, we shall consider the case (ii) in Theorem 1. We believe that this will bring interesting links with the partially unipolar bi-capacities.

## References

1. Von Neuman, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press (1944)
2. Savage, L.J.: The Foundations of Statistics, 2nd edn., Dover (1972)
3. Krantz, D., Luce, R., Suppes, P., Tversky, A.: Foundations of measurement. Additive and Polynomial Representations, vol. 1. Academic Press (1971)
4. Schmeidler, D.: Subjective probability and expected utility without additivity. Econometrica 57(3), 571–587 (1989)
5. Grabisch, M.: The application of fuzzy integrals in multicriteria decision making. European J. of Operational Research 89, 445–456 (1996)
6. Labreuche, C., Grabisch, M.: Generalized Choquet-like aggregation functions for handling ratio scales. European Journal of Operational Research 172, 931–955 (2006)
7. Labreuche, C.: Construction of a Choquet integral and the value functions without any commensurateness assumption in multi-criteria decision making. In: Int. Conf. Of the Euro Society for Fuzzy Logic and Technology (EUSFLAT), Aix Les Bains, France, July 18-22 (2011)
8. Labreuche, C., Grabisch, M.: The representation of conditional relative importance between criteria. Annals of Operation Research 154, 93–122 (2007)
9. Bana e Costa, C.A., Vansnick, J.C.: A theoretical framework for Measuring Attractiveness by a Categorical Based Evaluation TecHnique (MACBETH). In: Proc. XIth Int. Conf. on MultiCriteria Decision Making, Coimbra, Portugal, pp. 15–24 (August 1994)
10. Labreuche, C., Grabisch, M.: The Choquet integral for the aggregation of interval scales in multicriteria decision making. Fuzzy Sets & Systems 137, 11–26 (2003)
11. Choquet, G.: Theory of capacities. Annales de l'Institut Fourier 5, 131–295 (1953)
12. Grabisch, M., Labreuche, C.: Bi-capacities for decision making on bipolar scales. In: EUROFUSE Workshop on Informations Systems, Varenna, Italy, pp. 185–190 (September 2002)
13. Gonzales, C., Perny, P.: GAI networks for utility elicitation. In: Proceedings of the 9th International Conference on the Principles of Knowledge Representation and Reasoning, pp. 224–234 (2004)

# Multi-valued Representation of Neutrosophic Information

Vasile Patrascu

Tarom Information Technology, Bucharest, Romania
patrascu.v@gmail.com

**Abstract.** The paper presents three variants for multi-valued representation of neutrosophic information. These three representations are provided in the framework of multi-valued logics and it provides some calculation formulae for the following neutrosophic features: *truth, falsity, neutrality, undefinedness, saturation, contradiction, ambiguity.* In addition, it was defined *net-truth, definedness, neutrosophic score* and *neutrosophic indeterminacy*.

**Keywords:** Neutrosophic information, multi-valued logic, truth, falsity, neutrality, undefinedness, saturation, ambiguity, contradiction, neutrosophic indeterminacy.

## 1 Introduction

In the *fuzzy representation* of information proposed by Zadeh [19], the degree of truth is described by $T$ while the degree of falsity is the negation of $T$, namely $F = 1 - T$. We extend the representation of fuzzy information to a three-valued one by defining the truth, falsity and ambiguity [11]:

$$t = T - \min(T, F) \tag{1.1}$$

$$f = F - \min(F, T) \tag{1.2}$$

$$a = 2\min(T, F) \tag{1.3}$$

with the partition: $$t + a + f = 1 \tag{1.4}$$

Thus we have extracted three fuzzy information features and $t \cdot f = 0$. The ambiguity represents the only one component of indeterminacy, namely $i = a$.

In the *intuitionistic fuzzy representation* of information proposed by Atanassov [1], the degree of truth is described by $T$, the degree of falsity is described by $F$ with the condition $T + F \le 1$. We extend the representation of intuitionistic fuzzy information to a four-valued one defining the truth, falsity, ambiguity and undefinedness [11]:

$$t = T - \min(T, F) \tag{1.5}$$

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 304–313, 2014.

$$f = F - \min(F,T) \tag{1.6}$$

$$a = 2\min(T,F) \tag{1.7}$$

$$u = 1 - T - F \tag{1.8}$$

with the partition: $$t + a + u + f = 1 \tag{1.9}$$

Thus, we have extracted four fuzzy information features and $t \cdot f = 0$. The indeterminacy has two components: ambiguity and undefinedness, namely $i = a + u$.

The *bifuzzy information* is provided by the degree of truth $T$ and the degree of falsity $F$. Atanassov proposed a three-valued representation [2] defining the truth, falsity, and undefinedness:

$$t = T - \frac{\min(T,F)}{2} \tag{1.10}$$

$$f = F - \frac{\min(T,F)}{2} \tag{1.11}$$

$$u = 1 - \max(T,F) \tag{1.12}$$

with the partition: $$t + u + f = 1 \tag{1.13}$$

Also, for bifuzzy case, a penta-valued representation based on truth, falsity, undefinedness, contradiction and ambiguity was proposed in [9], [11] and [12]:

$$t = T - \min(T,F) \tag{1.14}$$

$$f = F - \min(F,T) \tag{1.15}$$

$$u = 1 - \min(T + F,1) \tag{1.16}$$

$$c = \max(T + F,1) - 1 \tag{1.17}$$

$$a = 1 - |T - F| - |T + F - 1| \tag{1.18}$$

with the fuzzy partition:

$$t + a + u + c + f = 1 \tag{1.19}$$

and having the properties: $u \cdot c = 0$, $t \cdot f = 0$. The indeterminacy has three components: ambiguity, undefinedness and contradiction, namely $i = a + u + c$.

The *neutrosophic representation* of information was proposed by Smarandache [15-18] and it is provided by the triplet (*T,I,F*): degree of truth, degree of indeterminacy and degree of falsity. Analyzing the components, we can observe, the bifuzzy pair has its own indeterminacy defined by similarity between $T$ and $F$, by example $I = 1 - |T - F|$. In this context, we have two values for indeterminacy: the first is explicit and the second is implicit. This fact could generate some confusions and the neutrosophic model could become inconsistent. By example, for (1,0,1) the explicit

indeterminacy is 0 but the bifuzzy information has a contradictory nature and its indeterminacy is 1. Because of that, in this paper we will consider that the neutrosophic information is obtained adding the neutrality $N$ to the bifuzzy information $(T,F)$. The triplet of neutrosophic information will be denoted with $(T,N,F)$. This paper will redefine for the neutrosophic information the features like truth, falsity, undefinedness, contradiction, ambiguity, indeterminacy, net truth, definedness, neutrosophic score and it introduces new features like neutrality and saturation. In the following, the paper has the structure: Section 2 presents the construction of net truth, definedness, neutrosophic score; Section 3 presents some calculation formulae for neutrosophic indeterminacy; Section 4 presents a tetra-valued representation based on truth, falsity, neutrality and undefinedness; Section 5 presents a penta-valued representation of neutrosophic information based on truth, falsity, neutrality, saturation and undefinedness; Section 6 presents a hepta-valued representation based on truth, falsity, neutrality, saturation, contradiction, ambiguity and undefinedness; Section 7 outlines some conclusions.

## 2 The Net-Truth, Definedness and Neutrosophic Score

We extend the notions of net truth and definedness from bifuzzy information to neutrospohic one. For bifuzzy information it was defined in [12] the net-truth $\tau$ and definedness $\delta$ .We will define the net-truth and definedness [14] in the following way:

$$\tau = \frac{T-F}{1+N} \tag{2.1}$$

$$\delta = \frac{3(T+N+F-1)-|T+N+F-1|}{4} \tag{2.2}$$

If the neutrosophic definedness is positive, the information is or over-defined, if it is zero, the neutrosophic information is complete and if it is negative, the information is incomplete. The pair $(\tau,\delta)$ provides a bi-valued representation of neutrosophic information. Combining these two components into a scalar, one obtains the neutrosophic score defined by:

$$\eta = \frac{\tau}{1+|\delta|} \tag{2.3}$$

The neutrosophic score defines the following order relation:

$$(T_1,I_1,F_1) \geq (T_2,I_2,F_2) \quad \Leftrightarrow \quad \eta_1 \geq \eta_2 \tag{2.4}$$

The proposed order relation (2.4) could be used in decision making procedures if the neutrosophic information is employed.

## 3 The Neutrosophic Indeterminacy

For neutrosophic indeterminacy, we will trace the Kosko idea for fuzziness calculation [6]. Kosko proposed to measure this information feature by a similarity function between the distance to the nearest crisp element and the distance to the farthest crisp element. For neutrosophic information the two crisp elements are $(1,0,0)$ and $(0,0,1)$ .

We consider the following vector: $V=(T-F,T+F-1,N)$. For $(1,0,0)$ and $(0,0,1)$ it results: $V_T=(1,0,0)$ and $V_F=(-1,0,0)$. We will compute the distances:

$$D(V,V_T)=|T-F-1|+|T+F-1|+N \tag{3.1}$$

$$D(V,V_F)=|T-F+1|+|T+F-1|+N \tag{3.2}$$

The indeterminacy will be defined by the similarity between these two distances.

Using the Czekanowskyi formula [4] it results the similarity and the indeterminacy:

$$S_C=1-\frac{|D(V,V_T)-D(V,V_F)|}{D(V,V_T)+D(V,V_F)} \tag{3.3}$$

$$I=1-\frac{|T-F|}{1+N+|T+F-1|} \tag{3.4}$$

For $N=0$ and $T+F\le 1$ it results the intuitionistic fuzzy entropy [13]. For $T+F=1$ and $N=0$, it results the fuzzy entropy proposed by Kaufman [5].

Using the Ruzicka formula [4] it results the similarity and indeterminacy formulae:

$$S_R=1-\frac{|D(V,V_T)-D(V,V_F)|}{\max\big(D(V,V_T),D(V,V_F)\big)} \tag{3.5}$$

$$I=1-\frac{2|T-F|}{1+|T-F|+N+|T+F-1|} \tag{3.6}$$

For $N=0$, it result the bifuzzy entropy [12]. For $T+F=1$ and $N=0$, it results the fuzzy entropy proposed by Kosko [6]. We notice that the neutrosophic indeterminacy is a strictly decreasing function in $|T-F|$ and non-decreasing in $N$ and in $|T+F-1|$. Taking into account these properties we can construct other formulae, by example:

$$I=1-\frac{|T-F|}{1+N} \tag{3.7}$$

$$I=1-\frac{|T-F|}{1+\max(N,|T+F-1|)} \tag{3.8}$$

$$I=1-\frac{|T-F|}{\sqrt{0.5+(T-0.5)^2+(F-0.5)^2+N^2}} \tag{3.9}$$

## 4 Tetra-Valued Representation of Neutrosophic Information

The three valued representation of bifuzzy information (1.10-1.12) will be extended for neutrosophic information adding the neutrality feature. Formulae (1.10-1.12) compress the unit square $[0,1]^2$ to a triangle with the vertices $B_T=(1,0)$, $B_F=(0,1)$ and $B_U=(0,0)$. Here, we will extend this compression from 2-dimensional space to 3-dimensional one. We will compress the unit cube $[0,1]^3$ (the primary space of

neutrosophic information) to a tetrahedron with the vertices $V_T = (1,0,0)$, $V_F = (0,0,1)$, $V_N = (0,1,0)$, $V_U = (0,0,0)$. We define the truth, falsity, neutrality and undefinedness [14]:

$$t = T - \frac{\min(T,N) + \min(T,F)}{2} + \frac{\min(T,N,F)}{3} \tag{4.1}$$

$$f = F - \frac{\min(F,N) + \min(T,F)}{2} + \frac{\min(T,N,F)}{3} \tag{4.2}$$

$$n = N - \frac{\min(T,N) + \min(N,F)}{2} + \frac{\min(T,N,F)}{3} \tag{4.3}$$

$$u = 1 - \max(T,N,F) \tag{4.4}$$

These four parameters define a partition of unity.

$$t + f + n + u = 1 \tag{4.5}$$

Having this representation, the neutrosophic information could be *true, false, neutral or undefined*. The *indeterminacy* has three components: *neutrality*, *undefinedness* and *ambiguity*:

$$i = n + u + a \tag{4.7}$$

where *ambiguity* is defined by:

$$a = 2\min(t, f) \tag{4.8}$$

The formulae (4.1 - 4.4) provide a way to transform any neutrosophic information into an incomplete one, because from (4.5) it results:

$$t + f + n \le 1 \tag{4.9}$$

We must mention that, for $N = 0$, one obtains (1.10), (1.11) and (1.12) proposed by Atanassov for transforming a bifuzzy information into an intuitionistic one [2].

We define the union, intersection and the complement operators.

- *The Complement*:

For any $q = (t,n,u,f)$ the complement is defined by:

$$\overline{q} = (f,n,u,t) \tag{4.10}$$

For the union and intersection we will use the formulae proposed in [8], [10]. The formulae proposed here are different from those proposed by Ashbacher [3].

- *The Union*

For $q_1 = (t_1,n_1,u_1,f_1)$ and $q_2 = (t_2,n_2,u_2,f_2)$

$$t_{q_1 \cup q_2} = t_1 \vee t_2 \tag{4.11}$$

$$n_{q_1 \cup q_2} = (n_1 + t_1) \vee (n_2 + t_2) - t_1 \vee t_2 \tag{4.12}$$

$$u_{q_1 \cup q_2} = (u_1 + f_1) \wedge (u_2 + f_2) - f_1 \wedge f_2 \tag{4.13}$$

$$f_{q_1 \cup q_2} = f_1 \wedge f_2 \tag{4.14}$$

- *The Intersection*:

$$t_{q_1 \cap q_2} = t_1 \wedge t_2 \tag{4.15}$$

$$u_{q_1 \cap q_2} = (u_1 + t_1) \wedge (u_2 + t_2) - t_1 \wedge t_2 \tag{4.16}$$

$$n_{q_1 \cap q_2} = (n_1 + f_1) \vee (n_2 + f_2) - f_1 \vee f_2 \tag{4.17}$$

$$f_{q_1 \cap q_2} = f_1 \vee f_2 \tag{4.18}$$

where $\wedge$ represents any Frank t-norm [7] and $\vee$ represents its t-conorm. The above remark is valid for all the next sections of this paper where the symbols $\wedge, \vee$ are used.

Finally, after union or intersection operations, we can obtain the values for the primary space, using the following inverse transform from the tetrahedron to the unit cube:

$$T = t + \min(t, n) + \min(t, f) \tag{4.19}$$

$$N = n + \min(n, t) + \min(n, f) \tag{4.20}$$

$$F = f + \min(f, t) + \min(f, n) \tag{4.21}$$

## 5 Penta-Valued Representation of Neutrosophic Information Based on Definedness

We will compress the unit cube $[0,1]^3$ to a double tetrahedron with the vertices $V_T = (1,0,0)$ , $V_F = (0,0,1)$ , $V_N = (0,1,0)$ , $V_S = (1,1,1)$ , $V_U = (0,0,0)$ . The triangle $(V_T, V_F, V_N)$ (truth-falsity-neutrality) defines de base, the saturation point $V_S$ is over the base while the undefinedness point $V_U$ is under the base. Using the neutrosophic definedness (2.2), we define the truth, falsity, neutrality, saturation and undefinedness by:

$$t = \frac{1 - \max(\delta, 0)}{T + N + F + \max(-\delta, 0)} \cdot T \tag{5.1}$$

$$n = \frac{1 - \max(\delta, 0)}{T + N + F + \max(-\delta, 0)} \cdot N \tag{5.2}$$

$$f = \frac{1 - \max(\delta, 0)}{T + N + F + \max(-\delta, 0)} \cdot F \tag{5.3}$$

$$s = \max(\delta, 0) \tag{5.4}$$

$$u = \frac{\max(-\delta, 0)}{T + N + F + \max(-\delta, 0)} \tag{5.5}$$

Due to the particular form of definedness (2.2), the saturation and undefinedness have the following equivalent forms:

$$\begin{cases} s = \max(T + F + N - 1, 0)/2 \\ u = \max(1 - T - F - N, 0) \end{cases} \tag{5.6}$$

These five parameters verify the condition of fuzzy partition, namely:

$$t + f + n + s + u = 1 \tag{5.7}$$

Having this representation, the neutrosophic information could be *true, false, neutral, saturated* and *undefined*. In the first step, for this representation the *indeterminacy* has four components: *neutrality*, *saturation*, *undefinedness* and *ambiguity*:

$$i = n + s + u + a \tag{5.8}$$

where *ambiguity* is defined by:

$$a = 2\min(t, f) \tag{5.9}$$

In the second step, we extract the *contradiction* $c$ from the pair $(s, a)$ by $c = 2\min(s, a)$ while saturation and ambiguity become $s^+ = c - \min(s, a)$, $a^+ = a - \min(s, a)$ and we obtain for indeterminacy a five-valued representation:

$$i = n + s^+ + u + c + a^+ \tag{5.10}$$

We define the union, intersection and the complement operators.

- *The Complement*:

For any $q = (t, n, s, u, f)$ the complement is defined by:

$$\overline{q} = (f, n, s, u, t) \tag{5.11}$$

For the union and intersection we will use the formulae proposed in [9].

- *The Union*

For $q_1 = (t_1, n_1, s_1, u_1, f_1)$ and $q_2 = (t_2, n_2, s_2, u_2, f_2)$

$$t_{q_1 \cup q_2} = t_1 \vee t_2 \tag{5.12}$$

$$s_{q_1 \cup q_2} = (s_1 + f_1) \wedge (s_2 + f_2) - f_1 \wedge f_2 \tag{5.13}$$

$$u_{q_1 \cup q_2} = (u_1 + f_1) \wedge (u_2 + f_2) - f_1 \wedge f_2 \tag{5.14}$$

$$f_{q_1 \cup q_2} = f_1 \wedge f_2 \tag{5.15}$$

$$n_{q_1 \cup q_2} = 1 - t_{q_1 \cup q_2} - f_{q_1 \cup q_2} - s_{q_1 \cup q_2} - u_{q_1 \cup q_2} \tag{5.16}$$

- *The Intersection*:

$$t_{q_1 \cap q_2} = t_1 \wedge t_2 \tag{5.17}$$

$$s_{q_1 \cap q_2} = (s_1 + t_1) \wedge (s_2 + t_2) - t_1 \wedge t_2 \tag{5.18}$$

$$u_{q_1 \cap q_2} = (u_1 + t_1) \wedge (u_2 + t_2) - t_1 \wedge t_2 \tag{5.19}$$

$$f_{q_1 \cap q_2} = f_1 \vee f_2 \tag{5.20}$$

$$n_{q_1 \cap q_2} = 1 - t_{q_1 \cap q_2} - f_{q_1 \cap q_2} - s_{q_1 \cap q_2} - u_{q_1 \cap q_2} \tag{5.21}$$

## 6 Hepta-Valued Representation of Neutrosophic Information Based on Indeterminacy Decomposition

The penta-valued representation of bifuzzy information (1.14-1.18) will be extended for neutrosophic information adding neutrality and saturation. Using the indeterminacy defined by (3.7) we will define a hepta-valued representation base on the following features: truth *t*, falsity *f*, ambiguity *a*, contradiction *c*, saturation *s*, neutrality *n* and undefinedness *u*. Firstly, we define the following auxiliary parameters:

$$C = \max(T + F,1) - 1 \tag{6.1}$$

$$U = 1 - \min(T + F,1) \tag{6.2}$$

Using the net-truth, we define de index of truth and the index of falsity:

$$t = \frac{T - \min(T,F)}{1+N} \tag{6.3}$$

$$f = \frac{F - \min(F,T)}{1+N} \tag{6.4}$$

We obtained the following three-valued fuzzy partition:

$$t + f + I = 1 \tag{6.5}$$

The indeterminacy will have the same structure with that defined by (5.10). In the next, we will identify the components of indeterminacy: ambiguity, contradiction, undefinedness, saturation and neutrality:

$$a = \frac{1 - |T - F| - U - C}{1+N} \tag{6.6}$$

$$c = \frac{C - \min(N,C)}{1+N} \tag{6.7}$$

$$u = \frac{U - \min(N,U)}{1+N} \tag{6.8}$$

$$s = \frac{N/2 - \min(N,U)/2 + 3\min(N,C)/2}{1+N} \tag{6.9}$$

$$n = \frac{N/2 - \min(N,C)/2 + 3\min(N,U)/2}{1+N} \tag{6.10}$$

with the following property:

$$I = a + c + n + s + u \tag{6.11}$$

Finally from (6.5) it results the following hepta-valued fuzy partition of unity:

$$t + f + a + c + n + s + u = 1 \tag{6.12}$$

Having this representation, the neutrosophic information could be *true, false, ambiguous, contradictory, neutral, saturated* and *undefined*. These seven information features have the following prototypes: $V_T = (1,0,0)$, $V_F = (0,0,1)$, $V_A = (1/2,0,1/2)$, $V_C = (1,0,1)$, $V_N = (0,1,0)$, $V_S = (1,1,1)$ and $V_U = (0,0,0)$. The first two prototypes are related to certainty while the last five prototypes are related to uncertainty.

We define the union, intersection and the complement.

- *The Complement*:

For any $q=(t,a,c,n,s,u,f)$ the complement is defined by:

$$\overline{q}=(f,a,c,n,s,u,t) \tag{6.13}$$

For the union and intersection we will use the formulae proposed in [9].

- *The Union*

For $q_1=(t_1,a_1,c_1,n_1,s_1,u_1,f_1)$ and $q_2=(t_2,a_2,c_2,n_2,s_2,u_2,f_2)$

$$t_{q_1\cup q_2}=t_1\vee t_2 \tag{6.14}$$

$$n_{q_1\cup q_2}=(n_1+f_1)\wedge(n_2+f_2)-f_1\wedge f_2$$

$$c_{q_1\cup q_2}=(c_1+f_1)\wedge(c_2+f_2)-f_1\wedge f_2$$

$$s_{q_1\cup q_2}=(s_1+f_1)\wedge(s_2+f_2)-f_1\wedge f_2 \tag{6.15}$$

$$u_{q_1\cup q_2}=(u_1+f_1)\wedge(u_2+f_2)-f_1\wedge f_2 \tag{6.16}$$

$$f_{q_1\cup q_2}=f_1\wedge f_2 \tag{6.17}$$

$$a_{q_1\cup q_2}=1-t_{q_1\cup q_2}-f_{q_1\cup q_2}-n_{q_1\cup q_2}-c_{q_1\cup q_2}-s_{q_1\cup q_2}-u_{q_1\cup q_2} \tag{6.18}$$

- *The Intersection*:

$$t_{q_1\cap q_2}=t_1\wedge t_2 \tag{6.19}$$

$$n_{q_1\cap q_2}=(n_1+t_1)\wedge(n_2+t_2)-t_1\wedge t_2 \tag{6.20}$$

$$c_{q_1\cap q_2}=(c_1+t_1)\wedge(c_2+t_2)-t_1\wedge t_2 \tag{6.21}$$

$$u_{q_1\cap q_2}=(u_1+t_1)\wedge(u_2+t_2)-t_1\wedge t_2 \tag{6.22}$$

$$s_{q_1\cap q_2}=(s_1+t_1)\wedge(s_2+t_2)-t_1\wedge t_2 \tag{6.23}$$

$$f_{q_1\cap q_2}=f_1\vee f_2 \tag{6.24}$$

$$a_{q_1\cap q_2}=1-t_{q_1\cap q_2}-f_{q_1\cap q_2}-n_{q_1\cap q_2}-c_{q_1\cap q_2}-s_{q_1\cap q_2}-u_{q_1\cap q_2} \tag{6.25}$$

The complement operators defined by (4.10), (5.11) and (6.13) are compatible with the complement defined by Ashbacher [3].

## 7 Conclusion

Three multi-valued representations of neutrosophic information are presented in this paper, mainly in order to model features of its certainty and uncertainty. All the proposed representations verify the condition of fuzzy partition and are accompanied by operators like complement, union and intersection. It was extended the concepts of certainty like truth and falsity, the concepts of uncertainty like undefinedness, ambiguity and contradiction. In addition, it was defined new concepts related to the particularity of neutrosophy like saturation, neutrality, neutrosophic score and neutrosophic indeterminacy.

The particularization of the obtained formulae leads to some formulae that already exist in the specialty literature and are related to bifuzzy information, intuitionistic fuzzy information and fuzzy one. This fact proves the effectiveness of our approach.

## References


1. Atanassov, K.: Intuitionistic Fuzzy sets. Fuzzy Sets and Systems 20, 87–96 (1986)
2. Atanassov, K.: Remark on a Property of the Intuitionistic Fuzzy Interpretation Triangle. Notes on Intuitionistic Fuzzy Sets 8, 34 (2002)
3. Ashbacher, C.: Introduction to Neutrosophic logic. American Research Press, Rehoboth (2002)
4. Deza, E., Deza, M.M.: Dictionary of Distances. Elsevier (2006)
5. Kaufmann, A.: Introduction to the Theory of Fuzzy Subsets. Academic Press, New York (1975)
6. Kosko, B.: Fuzzy entropy and conditioning. Information Sciences 40, 165–174 (1986)
7. Frank, M.J.: On the simultaneous associativity of f(x,y) and x+y-f(x,y). Aeq. Math. 19, 194–226 (1979)
8. Patrascu, V.: Penta-Valued Fuzzy Set. In: The IEEE International Conference on Fuzzy Systems (FUZZY-IEEE 2007), London, U.K, pp. 137–140 (2007)
9. Patrascu, V.: Bi-fuzziness, Incompleteness, Inconsistency, Truth and Falsity Based on Saturation and Ignorance Functions. A New Approach of Penta-Valued Knowledge Representation. In: The 23rd Midwest Artificial Intelligence and Cognitive Science Conference, MAICS 2012, Cincinnati, OH, USA, pp. 112–118 (2012)
10. Patrascu, V.: Rough Sets on Four-Valued Fuzzy Approximation Space. In: The IEEE International Conference on Fuzzy Systems (FUZZY-IEEE 2007), London, U.K., pp. 133–136 (2007)
11. Patrascu, V.: A New Penta-valued Logic Based Knowledge Representation. In: Proc. of the 12th Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2008), Malaga, Spain, pp. 17–22 (2008)
12. Patrascu, V.: Cardinality and entropy for bifuzzy set. In: Proc. of the Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2010), Dortmund, Germany, pp. 656–665 (2010)
13. Patrascu, V.: Fuzzy Membership Function Construction Based on Multi-Valued Evaluation, in vol. Uncertainty Modeling in Knowledge Engineering and Decision Making. In: Proceedings of the 10th International FLINS Conference, pp. 756–761. World Scientific Press (2012)
14. Patrascu, V.: Neutrosophic information in the framework of multi-valued representation. In: The 21th Conference on Applied and Industrial Mathematics, CAIM 2013, Bucharest, Romania, September 19-22 (2013)
15. Smarandache, F.: Neutrosophy. / Neutrosophic Probability, Set, and Logic. American Research Press, Rehoboth (1998)
16. Smarandache, F.: Definiton of neutrosophic logic - a generalization of the intuitionistic fuzzy logic. In: Proceedings of the Third Conference of the European Society for Fuzzy Logic and Technology, EUSFLAT 2003, Zittau, Germany, pp. 141–146 (2003)
17. Smarandache, F.: Generalization of the Intuitionistic Fuzzy Logic to the Neutrosophic Fuzzy Set. International Journal of Pure and Applied Mathematics 24(3), 287–297 (2005)
18. Shumann, A., Smarandache, F.: Neutrality and Many-Valued Logics. American Research Press, USA (2007)
19. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)

# Formalising Information Scoring in a Multivalued Logic Framework

Adrien Revault d'Allonnes[1] and Marie-Jeanne Lesot[2,3]

[1] Université Paris 8, EA 4383, LIASD, FR-93526, Saint-Denis, France
Allonnes@ai.univ-paris8.fr
[2] Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France
[3] CNRS, UMR 7606, LIP6, FR-75005, Paris, France
Marie-Jeanne.Lesot@lip6.fr

**Abstract.** This paper addresses the task of information scoring seen as measuring the degree of trust that can be invested in a piece of information. To this end, it proposes to model the trust building process as the sequential integration of relevant dimensions. It also proposes to formalise both the degree of trust and the process in an extended multivalued logic framework that distinguishes between an indifferent level and the actual impossibility to measure. To formalise the process, it proposes multivalued combination operators matching the desired behaviours.

**Keywords:** trust, information quality, many-valued logic, information scoring, trust building.

## 1 Introduction

Amongst the diverse facets of information scoring, trust holds a central role. Qualifying the degree of trust that can be put in a piece of information can mean either evaluating its certainty – i.e. the reality of the fact it reports [1–3] – or the extent to which the rater is convinced, based on the process by which he forms an opinion about a hitherto unknown piece of information [4, 5].

Existing models differ in the dimensions they take into account to assess the degree of trust but also in the formal paradigms they use to represent it. Among the existing criteria, some examples are reliability, competence, sincerity, intention of the source, credibility, understood as confirmation by other sources, or plausibility, defined as likelihood with respect to *a priori* knowledge of the piece of information [1–6].

Information scoring as evaluation of trust has been formalised using both symbolic and numeric frameworks to evaluate each dimension and combine them using various aggregation operators to yield the final degree of trust. The original military model uses two discrete graded scales, respectively measuring source reliability and information credibility [1]. The aggregation is simply a concatenation of the two resulting symbols. This leads to difficulty in comparisons, hindering the evaluation's legibility [7]. Other models present the degree of trust

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 314–324, 2014.

as a single numerical value, a coefficient between 0 and 1, in either possibility [3] or evidence [6, 8, 9] theory frameworks. In the former, for instance, the dimensions are evaluated using possibility distributions, their aggregation relies on discounting, conjunctive and reinforcement operators [3].

This paper proposes to formalise information scoring in an extended multivalued logic framework: first off, as in the numerical approaches mentioned above, a unique degree is returned; second, as in the symbolic approach, it is defined on a discrete graded scale, which is further clarified with linguistic labels specifying the degrees' meaning, for all measured quantities [5]. Indeed, multivalued logic offers formal tools and concepts [10] which allow these improvements of the model's legibility. The proposed extension of the multivalued framework arises in the introduction of an additional degree modelling the case where a dimension cannot be quantified, through lack of knowledge, thus distinguishing it from the case where the evaluation leads to a neutral, indifferent value.

Each dimension mentioned above is individually evaluated in this extended multivalued framework. The resulting level of trust is also measured in this framework by successive integration of the dimensions: this paper proposes to model the trust building process as the sequential projection of the activation along selected dimensions on the current level of trust. To achieve this, the required properties for information scoring process projection operators are discussed and corresponding functions are proposed.

The paper is organised as follows: Section 2 presents the trust building process as the sequential projection of dimensions on the level of trust. Section 3 presents the proposed extended multivalued logic framework that distinguishes between lack of knowledge and indifference. Sections 4 and 5 are dedicated to the two types of combination operators identified. The paper then concludes with some directions for future work.

## 2 A Model of the Trust Building Process

This section presents a general model of the process through which trust in a piece of information is constructed. The principle of this model lies in updating the trust value by the sequential projection of the dimensions which partake in its evaluation. In particular, this section focuses on the attributes which will be formalised in the following sections, such as the influence of dimension projection, leading to specifications for the projection operators, as well as the difference between unknown and neutral levels.

### 2.1 Trust Building as a Sequence of Projections

The model described in this paper dynamically measures trust as a score whose value shifts with the consecutive consideration of different dimensions. These factors are projected on the current score, resulting in an updated trust value: trust-building is modelled as the consecutive integration of answers to different questions. This principle is illustrated in Figure 1, discussed further below, where

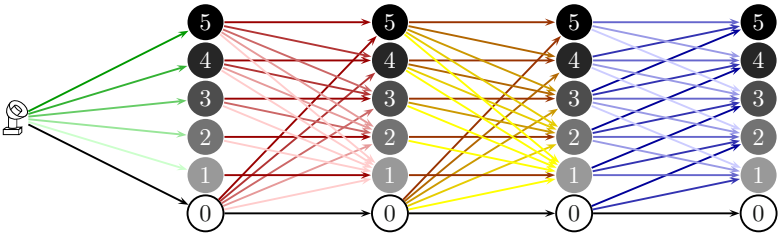



**Fig. 1.** Sequential projection of dimensions on trust

the shaded disks represent the current level of trust and the arrows four factors consecutively influencing its evolution. Obviously, the questions should not be redundant, nor should they be dependent. They should span different levels of specificity, from the most general, that is depending only on the context, not on the considered piece of information, to the particular, directly dependent on the informational content.

This principle is put in use in [5], where the dimensions are selected so as to consecutively answer the following questions: 'Who is telling me?', which corresponds to source reliability; 'What does he know about the topic?', to source competence; 'How likely does it seem?', to information plausibility; 'Is anyone else reporting it?', to information credibility. The first two relate to the source where the other two depend on the informational content. Furthermore, the evaluation of the source's reliability, how trustworthy the source is, is constant, whatever the considered piece of information, where competence depends both on the source and the topic, so is more specific to the piece of information. Plausibility, taken here as compatibility with the rater's knowledge, is contextual and subjective in that it depends both on the piece of information and the rater. Credibility is completely contextual, since it is taken as a degree of confirmation of the considered piece of information. The four dimensions therefore span the divide from source to information, general to contextual and subjective to objective, ensuring they are neither redundant nor dependent.

### 2.2 Combining the Dimensions: Influence of the Projections

Once the dimensions participating in the evaluation are selected, one must look at how their intensities are combined and the order in which they are considered. In order to model the trust building process, we propose to order the dimensions from the most immutable to the least, that is from the most general to the most specific, both content and contextwise. The idea behind this ordering is that the evaluation of trust starts with global dimensions and is progressively corrected with respect to the specifics of the considered piece of information.

These consecutive corrections entail that most projections have an abating influence. For instance, the model presented in [5] and briefly summarised above suggests that the evaluation starts with source reliability, the most general di-

mension as its value stays the same regardless of the content. It is then adapted to the assertion by the projection of source competence, this adaptation resulting in either a lower updated value or, at most, leaving it unchanged: faced with a trustworthy source, the fact that he is a specialist on the topic should not increase one's trust. Rather, if he were to talk about something which he knew nothing about, would one's doubts begin to rise. Similar arguments also apply to information plausibility. For this reason, an abating operator is needed to project most dimensions onto trust.

However, a second type of operator is necessary to model credibility, as this particular dimension can have both an increasing and weakening influence on the trust level. Indeed, the final step in a trust building process is to try to cross-check the informational content, i.e. to find confirmation or invalidation. Credibility appears in most existing models [1, 3, 5], usually, as the final projected dimension. This is in agreement with the specificity ordering proposed above, as cross-checking is both content and context dependent. Another argument for the credibility being the final step in the process is that its projection depends on the score of the confirmation which, thus, needs to have already been computed. Regarding credibility's influence, confirmations increase the trust level where invalidations weaken it, which explains the need for a specific cross-checking operator rather than the abating operator which, by definition, is one-way.

These two types of operators are illustrated in Figure 1, where the colour intensity of each arrow corresponds to the level to which each dimension is measured. The second and third sets of arrows, which all – except for the ones coming from the disks labelled 0, discussed in the following subsection – point downwards, are examples of abating operators. The final set of arrows, in both directions, is an example of the cross-checking operator.

### 2.3 Preserving the Expressiveness of Ignorance

The proposed sequential process relies on evaluations being made. However, there may be situations in which it is impossible to measure a particular dimension. It may, for instance, be impossible to evaluate an as yet unknown source's reliability, for lack of knowledge. Such a situation should not be confused with one in which a source is 'neither trustworthy nor untrustworthy', which conveys *some* knowledge on the matter. The latter is usually in the middleground of the evaluation scale. Distinguishing the two cases is necessary when projecting dimensions, as ignorance should not beget knowledge: if a dimension cannot be measured, its projection should not influence the outcome.

For this reason, a specific level is used in [1], representing the impossibility to measure. Likewise, in the example shown in Figure 1, this partly accounts for the horizontal arrows: projection of an impossible to measure dimension does not change the score, regardless of its current value. In addition, the disks labelled 0 represent a situation where the level of trust is unknown, i.e. none of the previous dimensions have been measured. In the absence of knowledge, the first dimension which can be evaluated moves the score out of ignorance. Note that, if evaluation

of trust was possible at any stage, it remains possible in all subsequent stages. Thus, no arrow points to a disk marked 0.

## 3 Extended Multivalued Logic Framework

The multivalued logic paradigm is based on a symbolic, discrete truth scale and offers a formal framework to manipulate and combine these degrees [10]. It constitutes a candidate framework to model the trust building process described in the previous section, addressing the clarity and legibility concerns: the discrete scales used both for measuring the dimensions and the level of trust, such as that presented in Figure 1, can be mapped to multivalued logic degrees. Moreover, multivalued logic combination operators can help express trust as a projection of dimensions in a single degree. However, to fully express all aspects of the model, the multivalued logic paradigm needs to be able to distinguish between an indifferent and an unknown degree.

After having recalled the essential characteristics of multivalued logic, this section describes the proposed extension and discusses its properties.

### 3.1 Multivalued Logic

Multivalued logic models reasoning using the $M$ truth degrees of a totally ordered set $\mathcal{L}_M = \{\tau_0, \ldots, \tau_{M-1}\}$, where $\tau_\alpha \leq \tau_\beta \Leftrightarrow \alpha \leq \beta$, coupled to an involutive negation satisfying De Morgan's laws. The ordering property guarantees all degrees in $\mathcal{L}_M$ are comparable. They span, at a granularity varying with $M$, the different levels of veracity they represent from $\tau_0$, meaning 'false', to $\tau_{M-1}$, for 'true'. The switch between the two appears around the middle value $\tau_{\frac{M-1}{2}}$, usually forced into the chosen scale by choosing an odd $M$. Furthermore, all $\tau_\alpha$'s come with a semantic label increasing their legibility.

On top of the expressiveness and legibility of the truth scale, multivalued logic offers tools to reason with truth degrees, through operators that generalise conjunction, disjunction or implication [10]. It also has operators for symbolic arithmetic, beyond a purely logic interpretation [11].

### 3.2 Introduction of an Ignorance Degree: $\tau_?$

Discrete scales used for information scoring can be mapped to multivalued logic degrees. In accordance with what was discussed in the previous section and in Section 2.3, the indifferent value can be mapped to $\tau_{\frac{M-1}{2}}$, leaving the question of representing the absence of knowledge.

We propose to extend the multivalued logic framework by introducing an additional degree, denoted $\tau_?$, in order to distinguish between ignorance and indifference. We propose to define this degree by the following properties, where $\odot$ denotes any multivalued operator:

1. $\tau_? \notin \mathcal{L}_M$
2. $\forall \tau_\alpha \in \mathcal{L}_M,\ \tau_? \odot \tau_\alpha = \tau_\alpha \odot \tau_? = \tau_\alpha$
3. $\neg \tau_? = \tau_?$
4. $\tau_? \odot \tau_? = \tau_?$

The first property sets $\tau_?$ apart from other degrees, not submitting it to the same ordering constraints as $\tau_\alpha \in \mathcal{L}_M$: an element whose truth cannot be evaluated cannot be compared to an element whose truth is known, no matter how true. Therefore, we propose to define $\tau_?$ as an exception to the total order rule and of a different nature than other degrees. The second property defines $\tau_?$ as a neutral element for all operators, in particular for both conjunctions and disjunctions, which implies it does not preserve the order on $\mathcal{L}_M$. The last two properties define the behaviour of $\tau_?$ when combined with itself.

It can be shown easily that $\tau_?$ preserves some essential properties of a De Morgan algebra, specifically the involution of negation and De Morgan's laws, but contradicts the axioms of identity, complement as well as the absorption law.

### 3.3 Characteristics of the Extended Multivalued Framework

We propose that truth be evaluated on $\mathcal{L}_M$ extended with $\tau_?$, which we write $\mathcal{L}_M^e = \mathcal{L}_M \cup \{\tau_?\}$. This section shows how it offers an essential gain in expressiveness, illustrating it in the information scoring application.

As already stated, $\tau_?$ is not comparable to any degree, which may seem contradictory with the sought qualities of truth degrees. However, the localised loss of comparability in $\mathcal{L}_M^e$ is not an issue as $\tau_?$ never needs be compared to any degree, and, in fact, satisfies an expectation: a fact whose truth is given as $\tau_?$ is neither more, nor less true than any other. For this same reason, the introduction of $\tau_?$ places the comparison between a 'half true' fact and another back in its rightful place in $\mathcal{L}_M$'s semantic hierarchy.

When constructing or modifying a logic framework, one should take heed of the impact on the consistency of the resulting system, i.e. the impossibility to infer a contradiction. In the case of multivalued logics, this property is already relaxed since the laws of excluded middle and noncontradiction do not hold. Now adding $\tau_?$ does not introduce inconsistency: because $\tau_?$ is defined as a neutral element for all combination operators, any inconsistency proved after its inclusion will persist should it be removed.

## 4 Multivalued Abating Combination Operator

The previous section defined the extended formal framework $\mathcal{L}_M^e$. This section and the next propose operators for manipulating degrees in $\mathcal{L}_M^e$ and, more specifically, operators exhibiting the properties detailed in Section 2 to model trust building. These operators are mappings $\mathcal{L}_M^e \times \mathcal{L}_M^e \longrightarrow \mathcal{L}_M^e$ with current score and evaluation of the dimension for inputs and updated score for output.

This section focuses on the abating operator whose necessity is detailed in Section 2. A formal description of the required properties is given before a corresponding function is defined.

### 4.1 Formal Description of the Required Behaviour

The properties needed for an abating operator $F : \mathcal{L}_M^e \times \mathcal{L}_M^e \longrightarrow \mathcal{L}_M^e$ are:

- $\forall \tau_\alpha, \tau_\beta \in \mathcal{L}_M \; F(\tau_\alpha, \tau_\beta) \leq \tau_\alpha$
- $\forall \tau_\alpha \in \mathcal{L}_M^e \qquad F(\tau_?, \tau_\alpha) = F(\tau_\alpha, \tau_?) = \tau_\alpha$
- $\forall \tau_\alpha, \tau_\beta, \tau_\gamma \in \mathcal{L}_M$ if $\tau_\alpha \leq \tau_\beta$, then $F(\tau_\alpha, \tau_\gamma) \leq F(\tau_\beta, \tau_\gamma)$ and $F(\tau_\gamma, \tau_\alpha) \leq F(\tau_\gamma, \tau_\beta)$

The first property guarantees the abating behaviour, by imposing the updated value is at most equal to the current value $\tau_\alpha$. The second property defines $\tau_?$ as a neutral element of $F$: when the current score is still unknown, the value of the projected dimension directly determines the updated score. Conversely, when the level of the dimension cannot be measured, the score is not updated. The final property ensures the operator is increasing in both its arguments. Indeed, the higher the current score, the higher the updated one, for any correction; reciprocally, for any given level of trust, the more the dimension is activated, the higher the updated score.

### 4.2 Proposed Operator

In order to satisfy the properties listed in the previous subsection, one could consider extending t-norms. Indeed, t-norms, in the regular multivalued case, satisfy the first and third properties, leaving only the special case of $\tau_?$. However, t-norms have other properties, such as commutativity and associativity, which are superfluous in this particular context and would lead to unwanted constraints, limiting the expressiveness of the model. Furthermore, even if these properties were accepted, the different nature of the arguments, current score vs. dimension level, makes them inconsistent.

We, therefore, propose to define a new operator, illustrated in Figure 2, based on a set of parameters $\kappa_\alpha^\gamma \in \mathcal{L}_M$ that define transition thresholds between a current score $\tau_\alpha$ and its updated value $\tau_\gamma$:

$$F(\tau_\alpha, \tau_\beta) = \begin{cases} \min\{\tau_\gamma \in \mathcal{L}_M | \tau_\beta \leqslant \kappa_\alpha^\gamma\} & \text{if } \tau_\alpha, \tau_\beta \in \mathcal{L}_M \\ \tau_\beta & \text{if } \tau_\alpha = \tau_? \\ \tau_\alpha & \text{if } \tau_\beta = \tau_? \end{cases}$$

The first line sets the conditions for the transition from $\tau_\alpha$ to $\tau_\gamma$. The other cases describe $\tau_?$ as a neutral element for both arguments.

Note that consistency with the proposed model imposes constraints on the parameters $\kappa_\alpha^\gamma$: to guarantee the function is increasing, they must be ordered, so that if $\tau_\gamma \leq \tau_\delta$, then $\kappa_\alpha^\gamma \leq \kappa_\alpha^\delta$. However, these ordering constraints on $\kappa_\alpha^\gamma$ are imposed for a given value of $\tau_\alpha$ but constraints of the form $\kappa_\alpha^\beta = \kappa_\alpha^\gamma \oplus \kappa_\gamma^\beta$, where $\tau_\gamma$ is an intermediate value between $\tau_\alpha$ and $\tau_\beta$ and $\oplus$ a suitable sum operator [11], are not required: going directly from $\tau_\alpha$ to $\tau_\beta$ need not be equivalent to successive transitions from $\tau_\alpha$ to $\tau_\gamma$ and from $\tau_\gamma$ to $\tau_\beta$.

## 5 Multivalued Cross-Checking Combination Operator

Two remarkable properties of credibility projection require a dedicated operator: first it can both weaken or increase the level of trust. Second, it involves an

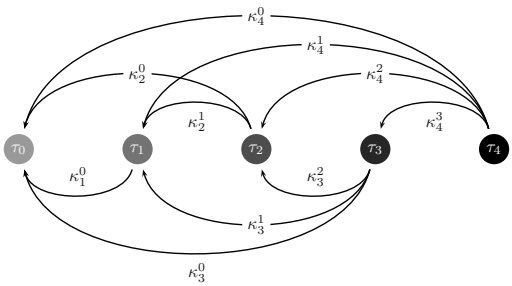



**Fig. 2.** An example of $F(\tau_\alpha, \tau_\beta)$ ignoring $\tau_?$: the degrees of $\mathcal{L}_5$ represent possible values of the current score, the arrows labelled with $\kappa_\alpha^\gamma$ transitions from $\tau_\alpha$ to $\tau_\gamma$

additional argument: besides the usual current score and dimension level, it depends on the current evaluation of the confirmation or invalidation. Indeed, an absolute confirmation should obviously increase the level of trust. However, if said corroboration has a low trust value, its influence should be less than that of a highly trusted confirmation.

### 5.1 Formal Description of the Required Behaviour

The cross-checking operator $G$ is therefore a mapping, $G : \mathcal{L}_M^e \times \mathcal{L}_M^e \times \mathcal{L}_M^e \longrightarrow \mathcal{L}_M^e$, satisfying the following properties:

- $G$ is increasing in all 3 arguments
- $\tau_?$ is a neutral element for all 3 arguments
- $\forall \tau_\alpha, \tau_\beta, \tau_\gamma \in \mathcal{L}_M \; G(\tau_\alpha, \tau_\beta, \tau_\gamma) \geq \tau_\alpha$ if $\tau_\beta \geq \tau_{\frac{M-1}{2}}$
  $G(\tau_\alpha, \tau_\beta, \tau_\gamma) \leq \tau_\alpha$ if $\tau_\beta < \tau_{\frac{M-1}{2}}$

The arguments for the monotonicity of operator $G$ are similar to the ones explaining $F$'s, as given in Section 4.1. The third property outlines the change of behaviour between confirmations (where credibility is in the top half of the scale: $\tau_\beta \geq \tau_{\frac{M-1}{2}}$), for which $G$ exhibits an increasing trend, and invalidations which have a weakening effect on the trust level.

### 5.2 Proposed Operator

Based on the properties required of the cross-checking operator, we propose the following function, illustrated in Figure 3:

$$G(\tau_\alpha, \tau_\beta, \tau_\gamma) = \begin{cases} \tau_\alpha & \text{if } \tau_\beta = \tau_? \\ \tilde{F}_i(\tau_\alpha, g(\tau_\beta, \tau_\gamma)) & \text{if } \tau_\beta < \tau_{\frac{M-1}{2}} \\ \tilde{F}_c(\tau_\alpha, g(\tau_\beta, \tau_\gamma)) & \text{otherwise} \end{cases}$$

As detailed below, $G$'s increasing and weakening trends are described by functions $\tilde{F}_c$ and $\tilde{F}_i$ respectively.

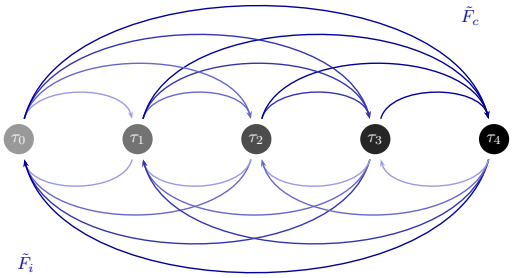



**Fig. 3.** An example of $G, \tau_\beta(\tau_\alpha, \tau_\gamma)$, in $\mathcal{L}_5$

In both cases, the range of influence is determined by a weighted credibility, defined as the aggregation of the degree of confirmation/invalidation, as measured by the credibility dimension $\tau_\beta$, and the trust level of the cross-checking piece of information $\tau_\gamma$. The aggregation operator $g : \mathcal{L}_M^e \times \mathcal{L}_M^e \longrightarrow \mathcal{L}_M^e$ must satisfy the following boundary conditions:

$$\forall \tau_\alpha, \tau_\beta \in \mathcal{L}_M \; g(\tau_\alpha, \tau_\beta) \leq \tau_\alpha \quad \text{if } \tau_\alpha \geq \tau_{\frac{M-1}{2}}$$
$$g(\tau_\alpha, \tau_\beta) \leq \neg\tau_\alpha \quad \text{if } \tau_\alpha < \tau_{\frac{M-1}{2}}$$

Indeed, the trust level of a confirmation can only weaken the impact of the credibility projection, along the same lines as the abating operator. The case of invalidation is similar, except for the negation of the credibility, which only serves to allow a symmetric behaviour. The monotonicity of $g$ must satisfy the following constraints, whose formal description is omitted for lack of space: $g$ must be increasing (resp. decreasing) in its first argument for confirmations (resp. invalidations); $g$ must be increasing in its second argument. Finally, $\tau_?$ must be an absorbing element for both arguments: if either cannot be evaluated, neither can the weighted credibility.

The $\tilde{F}_i$ operator, which controls the abating influence of weighted invalidations, is similar to the general abating operator defined in Section 4, its only specificity being its monotonicity in its second argument. Indeed, $\tilde{F}_i(\tau_\alpha,\tau_\beta)$ increases whereas $F(\tau_\alpha, \tau_\beta)$ decreases, when $\tau_\beta$ grows. Consequently, constraints on $\kappa_\alpha^\gamma$, and their comparisons with $\tau_\beta$, are reversed in $\tilde{F}_i$ defined similarly to $F$:

$$\tilde{F}_i(\tau_\alpha, \tau_\beta) = \begin{cases} \min\{\tau_\gamma \in \mathcal{L}_M | \tau_\beta \geq \kappa_\alpha^\gamma\} & \text{if } \tau_\alpha, \tau_\beta \in \mathcal{L}_M \\ \tau_\alpha & \text{if } \tau_\beta = \tau_? \\ \tau_\beta & \text{if } \tau_\alpha = \tau_? \end{cases}$$

An example of such an operator is represented in the bottom half of Figure 3.

The $\tilde{F}_c$ operator, which controls the emphasising influence of weighted confirmations, satisfies properties similar to that of $F$, except for the first one which

becomes $\forall \tau_\alpha, \tau_\beta \in \mathcal{L}_M, \tilde{F}_c(\tau_\alpha, \tau_\beta) \geq \tau_\alpha$, to express the emphasising influence. We thus propose to define:

$$\tilde{F}_c(\tau_\alpha, \tau_\beta) = \begin{cases} \max\{\tau_\gamma \in \mathcal{L}_M | \tau_\beta \geq \kappa_\alpha^\gamma\} & \text{if } \tau_\alpha, \tau_\beta \in \mathcal{L}_M \\ \tau_\alpha & \text{if } \tau_\beta = \tau_? \\ \tau_\beta & \text{if } \tau_\alpha = \tau_? \end{cases}$$

An example of such an operator is represented in the top half of Figure 3. Note that the definitions of the $\tilde{F}_i$ and $\tilde{F}_c$ functions are independent, it is not required that $\kappa_\alpha^\gamma = \kappa_\gamma^\alpha$ to allow for different sentivities to contradictory arguments.

## 6 Conclusion and Future Works

This paper proposes a sequential model for a trust building process, as the consecutive projection of dimensions on the current evaluation, as well as its formal and operational transposition in an extended multivalued logic framework. The latter offers a way to distinguish between indifference and ignorance, which is critical to adequately represent trust building. From the characterisation of the desired properties for the required projection operators, it defines special combination operators and gives the formal expression of suitable functions.

Future works will include the implementation of the proposed model and the definition of an evaluation protocol to lead an experimental study of user response to the proposed model. On a more formal level, another perspective lies in the theoretical study of the logical reasoning properties offered by the proposed extended multivalued logic $\mathcal{L}_M^e$, beyond the degree manipulation tools explored in this paper. Regarding information scoring, future works will aim at proposing tools allowing for the integration of the dynamics of trust building, like the influence on the evolution of trust of alternating confirmations and invalidations.

## References

1. NATO: Annex to STANAG2022. Information handling system (1997)
2. Demolombe, R.: Reasoning about trust: a formal logical framework. In: Jensen, C., Poslad, S., Dimitrakos, T. (eds.) iTrust 2004. LNCS, vol. 2995, pp. 291–303. Springer, Heidelberg (2004)
3. Lesot, M.J., Delavallade, T., Pichon, F., Akdag, H., Bouchon-Meunier, B.: Proposition of a semi-automatic possibilistic information scoring process. In: Proc. of EUSFLAT 2011, pp. 949–956 (2011)
4. Besombes, J., Revault d'Allonnes, A.: An extension of STANAG2022 for information scoring. In: Proc. of the Int. Conf. on Information Fusion, pp. 1635–1641 (2008)
5. Revault d'Allonnes, A.: An architecture for the evolution of trust: Definition and impact of the necessary dimensions of opinion making. In: Capet, P., Delavallade, T. (eds.) Information Evaluation, pp. 261–292. Wiley (2014)
6. Pichon, F., Dubois, D., Denœux, T.: Relevance and truthfulness in information correction and fusion. IJAR 53, 159–175 (2011)

7. Capet, P., Revault d'Allonnes, A.: Information evaluation in the military domain: Doctrines, practices and shortcomings. In: Capet, P., Delavallade, T. (eds.) Information Evaluation, pp. 103–128. Wiley (2014)
8. Cholvy, L.: Information evaluation in fusion: a case study. In: Proc. of IPMU 2004, pp. 993–1000 (2004)
9. Cholvy, L.: Evaluation of information reported: a model in the theory of evidence. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) IPMU 2010. CCIS, vol. 80, pp. 258–267. Springer, Heidelberg (2010)
10. Akdag, H., De Glas, M., Pacholczyk, D.: A qualitative theory of uncertainty. Fundamenta Informaticae 17, 333–362 (1992)
11. Seridi, H., Akdag, H.: Approximate Reasoning for Processing Uncertainty. J. of Advanced Comp. Intell. and Intell. Informatics 5, 110–118 (2001)

# Tableau Calculus for Basic Fuzzy Logic BL

Agnieszka Kułacka*

Department of Computing, Imperial College London
agnieszka.kulacka11@imperial.ac.uk
http://www.doc.ic.ac.uk/~ak5911/

**Abstract.** In this paper we present a tableau calculus for BL, basic fuzzy logic introduced by Petr Hájek in his monograph *Metamathematics of Fuzzy Logic*. We show that it is sound and complete with respect to continuous t-norms, and demonstrate the refutational procedure and the search for models procedure on a selected example. The idea of the calculus is based on the decomposition theorem for a continuous t-norm, by which this operation is shown to be equivalent to the ordinal sum of a family of t-norms defined on countably many intervals.

**Keywords:** tableaux, continuous t-norm, fuzzy logic.

## 1 Introduction

Basic fuzzy logic BL is a propositional logic in which the truth values of formulas are from interval $[0, 1]$, and was introduced by Petr Hájek in [7] as a generalisation of Łukasiewicz logic, Product logic and Gödel logic. Formulas of BL are written with propositional atoms and $\bar{0}$ (falsum) joined by strong conjunction $\&$ and implication $\rightarrow$. The other connectives $\wedge, \vee, \leftrightarrow$ and formula $\bar{1}$ (verum) may be treated as abbreviations. The semantics of the connectives $\&, \rightarrow$ are given by a continuous t-norm and its residuum. A t-norm is a function $\star : [0, 1]^2 \rightarrow [0, 1]$ that is associative, commutative and monotonic with $x \star 1 = x$ for $x \in [0, 1]$. Its residuum is $\Rightarrow$: $[0, 1]^2 \rightarrow [0, 1]$ defined by $x \Rightarrow y = \sup\{z : x \star z \leq y\}$ for $x, y \in [0, 1]$. Basic fuzzy logic BL has been proven in [3] and [6] to be the logic of continuous t-norms. Here, we take a further step and construct a tableau calculus for BL to prove validity of a formula or to construct a model if a formula is not valid.

A tableau calculus is a formal proof procedure that can be viewed as a refutational procedure to prove a given formula valid in a given logic, or as a search procedure for models with certain properties, in which a given formula is satisfiable, or both. In general, we start the procedure of tableau calculus with some expression, which asserts that a given formula $\psi$ is not valid. In the case of fuzzy logic BL, we will begin with an inequality $\Upsilon$ stating that the translated formula $\psi$ (to be described in full later in the paper) is less than 1. Then we apply some

* Supported by EPSRC DTA studentship in Computing Department, Imperial College London. I would like to thank Professor Ian Hodkinson and the reviewers for critically reading this paper and their constructive comments.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 325–334, 2014.

branch expansion rules, and either close all the branches using some closing rules, by which we fail to find a model that satisfies $\Upsilon$, or we obtain at least one open branch, by which we find a partial description of a model such that $\Upsilon$ is not contradictory in it. Thus, either we prove validity of $\psi$ by refutation, or we find a model in which $\psi$ is not valid.

Tableau methods for propositional logics have a long tradition (see [1], [5]) and have advantages over Hilbert systems such as bounded non-determinism, which means that there is a finite number of choices given by each branch expansion rule, and also the process of refutation and model searching terminates. They are automatic and thus well-suited for computer implementation. As described in the previous paragraph, not only can they show that a formula is valid, but in case it is not, they provide a model, in which the formula is not valid. There are some dual tableaux introduced for many-valued logics and fuzzy logics in [12] and [13]; however our approach is different. It follows in some respect the idea of the paper [8], where we constructed a relational hypersequents framework for Łukasiewicz modal fuzzy logic, but this approach is more general as far as the propositional rules are concerned since it deals with the logic of any continuous t-norm, not only that of Łukasiewicz t-norm. There are some approaches to construct proof systems based on a relational hypersequents frameworks for BL (see [2], [15]), though our approach involves inequalities that may be implemented with some adaptation of methodology to deal with operations on real numbers as in [16], and which exist due to occurence of Product proto-t-norms in some ordinal sums.

It should be noted that there are no fast and modern theorem provers and this limits real applications of fuzzy logics. There is also a gap for implementation of ordinal sums that contains Product proto-t-norms (see below). [16] We take a step towards filling this gap by introducing a tableau calculus for BL that is adaptable to logics that are complete with respect to any ordinal sums of Łukasiewicz and Product proto-t-norms.

The tableau calculus that is introduced in this paper is based on the decomposition theorem for continuous t-norms (see [4], [7], [9], [11]), which we will recall in section 2. After introducing the calculus in section 3, we will then demonstrate an application of the calculus. We prove soundness and completeness of the calculus in section 5 and we conclude the paper with some ideas for further research.

## 2 Decomposition Theorem

First, we will define Łukasiewicz and Product proto-t-norms and a special case of an ordinal sum of a family of continuous t-norms. This will be followed by the decomposition theorem for a continuous t-norm and the corollary for its residuum.

**Definition 1.** Let $0 \leq a^K < b^K \leq 1$ and $K \in \{L, P\}$.
1. *Łukasiewicz proto-t-norm defined on* $[a^L, b^L]$ is the function $\star_{a^L,b^L} : [a^L, b^L]^2 \to [a^L, b^L]$ given by equation (1). For every $x, y \in [a^L, b^L]$,

$$x \star_{a^L,b^L} y = \max\{a^L, x + y - b^L\}. \tag{1}$$

2. *Product proto-t-norm defined on* $[a^P, b^P]$ is the function $\star_{a^P,b^P} : [a^P, b^P]^2 \to [a^P, b^P]$ given by equation (2). For every $x, y \in [a^P, b^P]$,

$$x \star_{a^P,b^P} y = a^P + \frac{(x - a^P)(y - a^P)}{b^P - a^P}. \tag{2}$$

**Definition 2.** Let $\mathcal{I}$ be a countable index set and let $([a_n^{K_n}, b_n^{K_n}])_{n\in\mathcal{I}}$ be a family of intervals with $0 \le a_n^{K_n} < b_n^{K_n} \le 1, K_n \in \{L, P\}$ such that for all $i \ne j \in \mathcal{I}$, $(a_i^{K_i}, b_i^{K_i}) \cap (a_j^{K_j}, b_j^{K_j}) = \emptyset$. Then the *ordinal sum of a family of proto-t-norms* $(\star_{a_n^{K_n},b_n^{K_n}})_{n\in\mathcal{I}}$ is the function $\star : [0,1]^2 \to [0,1]$ defined by:

$$x \star y = \begin{cases} x \star_{a_n^{K_n},b_n^{K_n}} y & \text{if } x, y \in [a_n^{K_n}, b_n^{K_n}] \\ \min(x, y) & \text{otherwise,} \end{cases}$$

where $\star_{a_n^{K_n},b_{K_n}^{K}}$ is either Łukasiewicz proto-t-norm when $K_n = L$ or Product proto-t-norm when $K_n = P$ defined on $[a_n^{K_n}, b_n^{K_n}]$.

**Theorem 3.** (Decomposition theorem [4], [7], [9], [11])
The following two are equivalent:
1. The function $\star : [0,1]^2 \to [0,1]$ is a continuous t-norm.
2. The function $\star$ is the ordinal sum of a family of proto-t-norms $(\star_{a_n^{K_n},b_n^{K_n}})_{n\in\mathcal{I}}$, where $\mathcal{I}$ is a countable index set and $([a_n^{K_n}, b_n^{K_n}])_{n\in\mathcal{I}}$ is a family of intervals with $0 \le a_n^{K_n} < b_n^{K_n} \le 1, K_n \in \{L, P\}$ such that for all $i \ne j \in \mathcal{I}$, $(a_i^{K_i}, b_i^{K_i}) \cap (a_j^{K_j}, b_j^{K_j}) = \emptyset$.

**Corollary 4.** Let $\star$ be an ordinal sum of a family of proto-t-norms $(\star_{a_n^{K_n},b_n^{K_n}})_{n\in\mathcal{I}}$, where $\mathcal{I}$ is a countable index set and $([a_n^{K_n}, b_n^{K_n}])_{n\in\mathcal{I}}$ is a family of intervals with $0 \le a_n^{K_n} < b_n^{K_n} \le 1, K_n \in \{L, P\}$ such that for all $i \ne j \in \mathcal{I}$, $(a_i^{K_i}, b_i^{K_i}) \cap (a_j^{K_j}, b_j^{K_j}) = \emptyset$. Then, the residuum of $\star$ is the function $\Rightarrow : [0,1]^2 \to [0,1]$ given by equation (3). For every $x, y \in [0,1]$,

$$x \Rightarrow y = \begin{cases} 1 & \text{if } x \le y, \\ b_i^L - x + y & \text{if } x > y \text{ and } x, y \in [a_i^L, b_i^L], \\ a_i^P + \frac{(y - a_i^P)(b_i^P - a_i^P)}{x - a_i^P} & \text{if } x > y \text{ and } x, y \in [a_i^P, b_i^P], \\ y, & \text{otherwise,} \end{cases} \tag{3}$$

where $[a_i^L, b_i^L], [a_i^P, b_i^P]$ are intervals with Łukasiewicz proto-t-norm and Product proto-t-norm defined on them, respectively.

## 3 Tableau Calculus

In this section we will introduce a tableau calculus for showing that a formula $\psi$ of BL is valid or that there is a countermodel, in which the truth value of the formula $\psi$ is less than 1. First we define a tableau formula, then we define a translation function, by which we will get a term for a tableau formula at the root of a tree (also defined below) from a formula $\psi$ of BL.

**Definition 5.** (Tableau formula)
Let $L_0 = Par \cup \{+, -, \cdot, \div, \min, \max, \leq, <, 0, 1\}$ and $L_1 = L_0 \cup \{\star, \Rightarrow\}$ be signatures, where $Par$ is a set of constants (parameters), $+, -, \cdot, \div, \min, \max, \star, \Rightarrow$ are binary function symbols and $\leq, <$ are binary relation symbols. Let $Var$ be a set of variables.
1. If $x, y$ are $L_1$-terms, then $x \leq y, x < y, x = y$ are tableau formulas.
2. Let $E$ be a set of tableau formulas $s \leq t, s < t, s = t$, where $s, t$ are $L_0$-terms. We say that a mapping $\sigma : Var \to [0, 1]$ is a *solution* of $E$ iff there exists $L_0$-structure $\mathcal{M} = (\mathbb{R}, +, -, \cdot, \div, \min, \max, \leq, <, 0, 1, \rho)$, where $+, -, \cdot, \div$ (division), $\min, \max, 0, 1, \leq, <$ are interpreted as usual with $x \div 0$ assigned to 0 for any $x \in \mathbb{R}$ and $\rho : Par \to [0, 1]$ is a function, such that $\mathcal{M}, \sigma \models \bigwedge E$. We will call $\mathcal{M}$ an $L_0$-*structure modelling* $E$. By convention $\bigwedge \emptyset = \top$ (verum).[1]

**Definition 6.** (Translation function)
Let $\mathcal{F}$ be the set of formulas of BL and $\mathbb{T}$ be the set of $L_1$-terms. Let $\mu : PROP \to Var$ (we will write $\mu(p)$ as $\mu_p$) be a one-to-one mapping assigning variables to propositional atoms. Let $\psi, \varphi \in \mathcal{F}$. Then, we define a translation function $\tau : \mathcal{F} \to \mathbb{T}$, inductively:
1. $\tau(\bar{0}) = 0$, $\tau(\bar{1}) = 1$,
2. $\tau(p) = \mu(p)$ for every $p \in PROP$,
3. $\tau(\psi \& \varphi) = \tau(\psi) \star \tau(\varphi)$,
4. $\tau(\psi \to \varphi) = \tau(\psi) \Rightarrow \tau(\varphi)$,
5. $\tau(\psi \vee \varphi) = \max\{\tau(\psi), \tau(\varphi)\}$,
6. $\tau(\psi \wedge \varphi) = \min\{\tau(\psi), \tau(\varphi)\}$.

Let us recall some definitions from graph theory, which we will modify for the purposes of defining the calculus. A *graph* is a structure $(N, E)$, where $N$ is a finite set of nodes, $E$ is a set of edges such that $E \subseteq N \times N$ such that $\neg E(n, n)$ for all $n \in N$. A *successor* of $n \in N$ is $n' \in N$ iff there is an edge $e$ such that $(n, n') = e$. A *predecessor* of $n \in N$ is $n' \in N$ iff there is an edge $e'$ such that $(n', n) = e'$. A *path* from $n \in N$ to $n' \in N$ is a sequence of nodes $n_0 = n, n_1, ..., n_k = n'$ such that $k \geq 0$ and $n_i$ is a predecessor of $n_{i+1}$ for all $0 \leq i \leq k - 1$. A *leaf* is a node with no successors, and a *root* is a node with no predecessors. A *branch* is a path from a root to a leaf. We will call a *tree* an acyclic connected graph $(N, E)$, in which there is exactly one root and if a node is not a root, then it has exactly one predecessor. The height of a node $n \in N$ within a branch $\mathcal{B}$, denoted by $h(n, \mathcal{B})$, is the number of nodes on the path from $n$ to a leaf.

**Definition 7.** A *tableau* $\mathcal{T}$ for a formula $\psi$ of BL is a tree whose nodes are sets of tableau formulas and whose root is $\{\tau(\psi) < 1\}$, and on which the branch expansion rules have been fully applied. The rules of branch expansion are defined below, where $x, y, 0, 1$ are $L_1$-terms, $T$ is an $L_0$-term and $\Gamma$ is a set of tableau formulas. The multiple inequality should be understood in the usual way, e.g. instead of writing $a \leq c, c = d, d < f$, we write $a \leq c = d < f$. Let $\Diamond \in \{\leq, <\}$. Let $K, K_0, ..., K_{n-1} \in \{L, P\}$ be the labels as shown in the branch expansion rules. Suppose that parameters $a_0^{K_0} < b_0^{K_0} \leq a_1^{K_1} < ... \leq a_{n-1}^{K_{n-1}} < b_{n-1}^{K_{n-1}}$ $(n \geq 1)$ have been selected in the previous steps.

[1] Note that $\bigwedge$ is a symbol interpreted as conjunction in two-valued logic.

Let $a^K, b^K \in Par - \{a_i^{K_i}, b_i^{K_i} | 0 \leq i < n-1\}$ be distinct. Then $\mathcal{I}^K$ is exactly one of the sets:
Case 1. $\{0 \leq a^K < b^K \leq a_0^{K_0}\}$,
Case 2. $\{a^K = a_i^{K_i} < b^K = b_i^{K_i}\}$ for some $0 \leq i \leq n-1$ such that $K = K_i$,
Case 3. $\{b_i^{K_i} \leq a^K < b^K \leq a_{i+1}^{K_{i+1}}\}$ for some $0 \leq i \leq n-2$,
Case 4. $\{b_{n-1}^{K_{n-1}} \leq a^K < b^K \leq 1\}$.
If no parameters have been selected in the previous steps, then $\mathcal{I}^K$ is
Case 5. $\{0 \leq a^K < b^K \leq 1\}$.

The motivation for the following conditions is to take care of the cases when two $L_1$-terms $x, y$ (variables in the active term, i.e.one undergoing substitution, see below) are not in the same interval $(a_i^{K_i}, b_i^{K_i})$ (unlike in the cases above), which is equivalent to either $x$ is not in any such interval (cases 6, 9, 10) or $x$ is in such an interval and $y$ is not in the same one (the other cases). $\mathcal{J}$ is exactly one of the sets:
Case 6. $\{0 \leq x \leq a_0^{K_0}\}$,
Case 7. $\{a_i^{K_i} \leq x \leq b_i^{K_i}, y \leq a_i^{K_i}\}$ for some $0 \leq i \leq n-1$,
Case 8. $\{a_i^{K_i} \leq x \leq b_i^{K_i}, b_i^{K_i} \leq y\}$ for some $0 \leq i \leq n-1$,
Case 9. $\{b_i^{K_i} \leq x \leq a_{i+1}^{K_{i+1}}\}$ for some $0 \leq i \leq n-2$,
Case 10. $\{b_{n-1}^{K_{n-1}} \leq x \leq 1\}$.
If no parameters have been selected, then $\mathcal{J}$ is empty (Case 11.).

**Branch Expansion Rules**. Let $z_1, z_2$ be $L_1$-terms that contain $x \star y, x \Rightarrow y$ as a subterm, respectively. We will use a notation $z_1[v/x \star y]$ and $z_2[v/x \Rightarrow y]$ to denote that in $z_1, z_2$ we substituted an $L_1$-term $v$ for $x \star y$ and $x \Rightarrow y$, respectively, as shown in the rules below. We choose which subterm is an active term and assume that the tableau formula containing it does not belong to $\Gamma$.
**Rule** $(\star L)$**:** A branch with a node $\Gamma \cup \{z_1 \ \Diamond \ T\}$ expands following the subrules:
**Ł.** $\Gamma \cup \mathcal{I}^L \cup \{a^L \leq x \leq b^L, a^L \leq y \leq b^L, z_1[\max\{a^L, x+y-b^L\}/x \star y] \ \Diamond \ T\}$
**P.** $\Gamma \cup \mathcal{I}^P \cup \{a^P \leq x \leq b^P, a^P \leq y \leq b^P, z_1[a^P + \frac{(x-a^P)(y-a^P)}{b^P-a^P}/x \star y] \ \Diamond \ T\}$
**min.** $\Gamma \cup \mathcal{J} \cup \{z_1[\min\{x,y\}/x \star y] \ \Diamond \ T\}$
**Rule** $(\star R)$**:** A branch with a node $\Gamma \cup \{T \ \Diamond \ z_1\}$ expands following the subrules:
**Ł.** $\Gamma \cup \mathcal{I}^L \cup \{a^L \leq x \leq b^L, a^L \leq y \leq b^L, T \ \Diamond \ z_1[\max\{a^L, x+y-b^L\}/x \star y]\}$
**P.** $\Gamma \cup \mathcal{I}^P \cup \{a^P \leq x \leq b^P, a^P \leq y \leq b^P, T \ \Diamond \ z_1[a^P + \frac{(x-a^P)(y-a^P)}{b^P-a^P}/x \star y]\}$
**min.** $\Gamma \cup \mathcal{J} \cup \{T \ \Diamond \ z_1[\min\{x,y\}/x \star y])\}$
**Rule** $(\Rightarrow L)$**:** A branch with a node $\Gamma \cup \{z_2 \ \Diamond \ T\}$ expands following the subrules:
**All.** $\Gamma \cup \{x \leq y, z_2[1/x \Rightarrow y] \ \Diamond \ T\}$
**Ł.** $\Gamma \cup \mathcal{I}^L \cup \{a^L \leq y < x \leq b^L, z_2[b^L - x + y/x \Rightarrow y] \ \Diamond \ T\}$
**P.** $\Gamma \cup \mathcal{I}^P \cup \{a^P \leq y < x \leq b^P, z_2[a^P + \frac{(y-a^P)(b^P-a^P)}{x-a^P}/x \Rightarrow y] \ \Diamond \ T\}$
**min.** $\Gamma \cup \mathcal{J} \cup \{y < x, z_2[y/x \Rightarrow y] \ \Diamond \ T\}$
**Rule** $(\Rightarrow R)$**:** A branch with a node $\Gamma \cup \{T \ \Diamond \ z_2\}$ expands following the subrules:
**All.** $\Gamma \cup \{x \leq y, T \ \Diamond \ z_2[1/x \Rightarrow y]\}$
**Ł.** $\Gamma \cup \mathcal{I}^L \cup \{a^L \leq y < x \leq b^L, T \ \Diamond \ z_2[b^L - x + y/x \Rightarrow y]\}$
**P.** $\Gamma \cup \mathcal{I}^P \cup \{a^P \leq y < x \leq b^P, T \ \Diamond \ z_2[a^P + \frac{(y-a^P)(b^P-a^P)}{x-a^P}/x \Rightarrow y]\}$
**min.** $\Gamma \cup \mathcal{J} \cup \{y < x, T \ \Diamond \ z_2[y/x \Rightarrow y]\}$

Note that at node $n$, each of the branch expansion rules applied generates several nodes, which exact number depends on the number of parameters already at $n$ as more of the cases 1-11 will be used. For rules $(\star L), (\star R)$ we have subrules Ł., P., min., and for rules $(\Rightarrow L), (\Rightarrow R)$ additionally a subrule All. For the subrules Ł., P. of all rules, we have cases 1-5, and for the subrule min., we have cases 6-11.

The branch expansion rules are motivated by decomposition theorem 3 and its corollary 4.

**Definition 8.** For each branch $\mathcal{B}$ of a tableau $\mathcal{T}$ and each node $n \in \mathcal{B}$, we consider the set of $L_0$-formulas in $n$, $n\restriction_{L_0}$. We say that $\mathcal{B}$ is *closed* if for some node $n \in \mathcal{B}$, $n\restriction_{L_0}$ has no solution, otherwise it is *open*. Tableau $\mathcal{T}$ is *closed* if it only contains closed branches.[2] A tableau $\mathcal{T}$ is *open* if it has an open branch.

## 4 Example

Now we will show a tableau constructed for the prelinearity axiom with propositional atoms to illustrate the applicability of the rules. Note that this is only a part of a tableau, the branches passing through the nodes (11), (13) and (14) are missing, but they are easy to reconstruct.

**Example 9.** Let us build a tableau for formula $(p \to q) \vee (q \to p)$.

| | | | |
|---|---|---|---|
| **(1)** | $\{\max\{\mu_p \Rightarrow \mu_q, \mu_q \Rightarrow \mu_p\} < 1\}$ | | |
| **(11)** | $\{\mu_p \leq \mu_q, \max\{1, \mu_q \Rightarrow \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ All. | (to be completed) |
| **(12)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, \max\{b_0^L - \mu_p + \mu_q, \mu_q \Rightarrow \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ Ł. Case 5 | |
| **(12.1)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, \mu_q \leq \mu_p, \max\{b_0^L - \mu_p + \mu_q, 1\} < 1\}$ | Rule $(\Rightarrow L)$ All. | closed |
| **(12.2)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, 0 \leq a_1^L < b_1^L \leq a_0^L, a_1^L \leq \mu_p < \mu_q \leq b_1^L, \max\{b_0^L - \mu_p + \mu_q, b_1^L - \mu_q + \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ Ł. Case 1 | closed |
| **(12.3)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, a_0^L = a_1^L < b_1^L = b_0^L, a_1^L \leq \mu_p < \mu_q \leq b_1^L, \max\{b_0^L - \mu_p + \mu_q, b_1^L - \mu_q + \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ Ł. Case 2 | closed |
| **(12.4)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, b_0^L \leq a_1^L < b_1^L \leq 1, a_1^L \leq \mu_p < \mu_q \leq b_1^L, \max\{b_0^L - \mu_p + \mu_q, b_1^L - \mu_q + \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ Ł. Case 4 | closed |
| **(12.5)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, 0 \leq a_1^P < b_1^P \leq a_0^L, a_1^P \leq \mu_p < \mu_q \leq b_1^P, \max\{b_0^L - \mu_p + \mu_q, a_1^P + \frac{(\mu_p - a_1^P)(b_1^P - a_1^P)}{\mu_q - a_1^P}\} < 1\}$ | Rule $(\Rightarrow L)$ P. Case 1 | closed |
| **(12.6)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, b_0^L \leq a_1^P < b_1^P \leq 1, a_1^P \leq \mu_p < \mu_q \leq b_1^P, \max\{b_0^L - \mu_p + \mu_q, a_1^P + \frac{(\mu_p - a_1^P)(b_1^P - a_1^P)}{\mu_q - a_1^P}\} < 1\}$ | Rule $(\Rightarrow L)$ P. Case 4 | closed |
| **(12.7)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, 0 \leq \mu_q \leq a_0^L, \mu_p < \mu_q, \max\{b_0^L - \mu_p + \mu_q, \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ min. Case 6 | closed |
| **(12.8)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, a_0^L \leq \mu_q \leq b_0^L, \mu_p \leq a_0^L, \mu_p < \mu_q, \max\{b_0^L - \mu_p + \mu_q, \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ min. Case 7 | closed |
| **(12.9)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, a_0^L \leq \mu_q \leq b_0^L, b_0^L \leq \mu_p, \mu_p < \mu_q, \max\{b_0^L - \mu_p + \mu_q, \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ min. Case 8 | closed |
| **(12.10)** | $\{0 \leq a_0^L < b_0^L \leq 1, a_0^L \leq \mu_q < \mu_p \leq b_0^L, b_0^L \leq \mu_q \leq 1, \mu_p < \mu_q, \max\{b_0^L - \mu_p + \mu_q, \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ min. Case 10 | closed |
| **(13)** | $\{0 \leq a_0^P < b_0^P \leq 1, a_0^P \leq \mu_q < \mu_p \leq b_0^P, \max\{a_0^P + \frac{(\mu_q - a_0^P)(b_0^P - a_0^P)}{\mu_p - a_0^P}, \mu_q \Rightarrow \mu_p\} < 1\}$ | Case $(\Rightarrow L)$ P. Case 5 | (to be completed) |
| **(14)** | $\{\mu_q < \mu_p, \max\{\mu_q, \mu_q \Rightarrow \mu_p\} < 1\}$ | Rule $(\Rightarrow L)$ min. Case 11 | (to be completed) |

This tableau is closed as all its branches are closed (once it is completed).

[2] By Tarski theorem [14] on decidability of the first-order theory of $(\mathbb{R}, +, \cdot)$, it is decidable whether a tableau is closed.

## 5 Soundness and Completeness of Tableau Calculus

In the proof of soundness and completeness theorem of the tableau calculus, we will use a lemma showing the equivalence between existence of an open branch of a tableau with root $\{\tau(\psi) < 1\}$ and existence of a model, in which the formula $\psi$ is not valid. This, however, needs to be preceded by a definition of satisfiability of a branch.

**Definition 10.** Let $\mathcal{B}$ be a branch of tableau $\mathcal{T}$. Let $0 \leq a_0^{K_0} < b_0^{K_0} \leq ... \leq a_{n-1}^{K_{n-1}} < b_{n-1}^{K_{n-1}} \leq 1$, where $n \geq 0$ and $K_0, ..., K_{n-1} \in \{L, P\}$, be the inequalities between the parameters introduced by branch extension rules in $\mathcal{B}$. Let $l_{\mathcal{B}}$ be the leaf of the branch. Suppose that $\mathcal{M}$ is an $L_0$-structure modelling $l_{\mathcal{B}}$ and $\sigma : Var \to [0,1]$ such that $\mathcal{M}, \sigma \models \bigwedge l_{\mathcal{B}}$.
1. We expand the model $\mathcal{M}$ to an $L_1$-structure $\mathcal{M}_{\mathcal{B}} = (\mathbb{R}, +, -, \cdot, \div, \min, \max, \leq, <, 0, 1, \rho, \star_{\mathcal{B}}, \Rightarrow_{\mathcal{B}})$ such that for every $v, w \in [0,1]$,

$$v \star_{\mathcal{B}} w = \begin{cases} \max\{\rho(a_k^L), v + w - \rho(b_k^L)\} & \text{if } v, w \in [\rho(a_k^L), \rho(b_k^L)], 0 \leq k \leq n-1, \\ \rho(a_k^P) + \frac{(v - \rho(a_k^P)) \cdot (w - \rho(a_k^P))}{\rho(b_k^P) - \rho(a_k^P)} & \text{if } v, w \in [\rho(a_k^P), \rho(b_k^P)], 0 \leq k \leq n-1, \\ \min(v, w) & \text{otherwise,} \end{cases}$$

$$v \Rightarrow_{\mathcal{B}} w = \begin{cases} 1 & \text{if } v \leq w \\ \rho(b_k^L) - v + w & \text{if } \rho(a_k^L) \leq w < v \leq \rho(b_k^L), 0 \leq k \leq n-1, \\ \rho(a_k^P) + \frac{(w - \rho(a_k^P)) \cdot (\rho(b_k^P) - \rho(a_k^P))}{v - \rho(a_k^P)} & \text{if } \rho(a_k^P) \leq w < v \leq \rho(b_k^P), 0 \leq k \leq n-1, \\ w & \text{otherwise.} \end{cases}$$

2. A subset of a node $S$ of branch $\mathcal{B}$ is *$\mathcal{B}$-satisfiable via $\mathcal{M}, \sigma$* iff $\mathcal{M}_{\mathcal{B}}, \sigma \models \bigwedge S$, where the $L_1$-structure $\mathcal{M}_{\mathcal{B}}$ is constructed from $\mathcal{M}$ as in 1.
Now, we say that branch $\mathcal{B}$ is *satisfiable* iff there exist $\mathcal{M}, \sigma$ such that all nodes of $\mathcal{B}$ are $\mathcal{B}$-satisfiable via $\mathcal{M}, \sigma$.

**Lemma 11.** Let $\mathcal{T}$ be a tableau with a root $\{\tau(\psi) < 1\}$. Then the following are equivalent:
1. $\mathcal{T}$ has an open branch.
2. There is a model $\mathcal{A} = ([0,1], \star, \Rightarrow, 0, 1, V)$, where $\star$ is a continuous t-norm and $\Rightarrow$ is its residuum, $V : PROP \to [0,1]$ such that $V_\star(\psi) < 1$.[3]

*Proof.* Suppose that branch $\mathcal{B}$ of tableau $\mathcal{T}$ is open. So there is an $L_0$-structure $\mathcal{M}$ modelling leaf $l_{\mathcal{B}}$ of $\mathcal{B}$ and an assignment $\sigma : Var \to [0,1]$ such that $\mathcal{M}, \sigma \models l_{\mathcal{B}}$. We will construct a model $\mathcal{A}$ such that $V_\star(\psi) < 1$. First, we put $V(p) = \sigma(\tau(p))$ for all $p \in PROP$. We define operation $\star : [0,1]^2 \to [0,1]$ as $\star_{\mathcal{B}}$ and operation $\Rightarrow$: $[0,1]^2 \to [0,1]$ as $\Rightarrow_{\mathcal{B}}$ (see definition 10). By theorem 3 and corollary 4, $\star$ is a continuous t-norm with residuum $\Rightarrow$ and also by definitions 6 and 10, $V_\star(\psi) = [\![\tau(\psi)]\!]^{\mathcal{M}_{\mathcal{B}}, \sigma}$, where $[\![z]\!]^{\mathcal{M}_{\mathcal{B}}, \sigma}$ is the value of $L_1$-term $z$ in $\mathcal{M}_{\mathcal{B}}$ under the assignment $\sigma$. By induction on $h(m_{\mathcal{B}}, \mathcal{B})$, we show the claim that every

[3] $V_\star$ is the extension of $V$ to all formulas in BL with $V_\star(\bar{0}) = 0, V_\star(\bar{1}) = 1, V_\star(\psi \vee \varphi) = \max\{V_\star(\psi), V_\star(\varphi)\}, V_\star(\psi \wedge \varphi) = \min\{V_\star(\psi), V_\star(\varphi)\}, V_\star(\psi \& \varphi) = V_\star(\psi) \star V_\star(\varphi)$ and $V_\star(\psi \to \varphi) = V_\star(\psi) \Rightarrow V_\star(\varphi)$.

node $m_{\mathcal{B}}$ of $\mathcal{B}$ is $\mathcal{B}$-satisfiable via $\mathcal{M}, \sigma$. The sketch of the proof is as follows. Let $m'_{\mathcal{B}}$ be the node of $\mathcal{B}$ such that $h(m'_{\mathcal{B}}, \mathcal{B}) = h(m_{\mathcal{B}}, \mathcal{B}) + 1$, where $m_{\mathcal{B}}$ is assumed to be $\mathcal{B}$-satisfiable via $\mathcal{M}, \sigma$. Then $m'_{\mathcal{B}}$ is also $\mathcal{B}$-satisfiable via $\mathcal{M}, \sigma$ by definition 7. Therefore, in particular by this claim, the root of $\mathcal{T}$ is $\mathcal{B}$-satisfiable via $\mathcal{M}, \sigma$, and thus by definitions 6 and 10, $[\![\tau(\psi)]\!]^{\mathcal{M}_{\mathcal{B}}, \sigma} = V_\star(\psi) < 1$.

Conversely, suppose that there is a model $\mathcal{A} = ([0,1], \star, \Rightarrow, 0, 1, V)$, where $\star$ is a continuous t-norm and $\Rightarrow$ is its residuum, $V : PROP \to [0,1]$ such that $V_\star(\psi) < 1$. We know that for every node $n_{\mathcal{B}}$ of a branch $\mathcal{B}$, we have $n_{\mathcal{B}}{\restriction}_{L_0} \subseteq l_{\mathcal{B}}$, where $l_{\mathcal{B}}$ is the leaf of the branch. Therefore, a branch is open if its leaf $l_{\mathcal{B}}$ has a solution. That is, we need to find an $L_0$-structure modelling $l_{\mathcal{B}}$, say $\mathcal{M}$, and a mapping $\sigma : Var \to [0,1]$ such that $\mathcal{M}, \sigma \models \bigwedge l_{\mathcal{B}}$. First, we take $\sigma(\tau(p)) = V(p)$ for all $p \in PROP$. Now in (1) below, we construct the structure $\mathcal{M}$; the process reduces to constructing the mapping $\rho : PAR \to [0,1]$. At the same time we will be selecting nodes on a branch, we will call it $\mathcal{B}$. Then, in (2) we show that branch $\mathcal{B}$ is open.

(1) By theorem 3, $\star$ is defined as the ordinal sum of proto-t-norms $(\star_{\alpha_n^{K_n}, \beta_n^{K_n}})_{n \in \mathcal{I}}$. We will assign values of parameters occurring on $\mathcal{B}$ under $\rho$ to elements of $\{\alpha_n^{K_n}, \beta_n^{K_n} : n \in \mathcal{I}\}$. Suppose we selected the sequence of nodes $n_1, ..., n_l$, where $l \geq 1$, $n_1$ is the root of $\mathcal{T}$ and $n_{i+1}$ is the successor of $n_i$ for all $1 \leq i \leq l$. That is branch $\mathcal{B}$ is partially defined and $\rho$ is defined for all parameters on these nodes. By comparing $n_l$ and its successors, we deduce what the active term is and thus which of the branch expansion rules has been applied. Suppose it is $(\star L)$, the other cases are similar. Thus, there are subformulas of $\psi$, $\psi_1, \psi_2$, and $L_0$-term $T$,[4] such that $z \Diamond T$, where $\tau(\psi_1) \star \tau(\psi_2)$ is a subterm of $z$, belongs to $n_l$, but to none of its successors. We know whether or not there is $i \in \mathcal{I}$ such that $V_\star(\psi_1), V_\star(\psi_2) \in [\alpha_i^{K_i}, \beta_i^{K_i}]$. If there is such $i$ and $K_i = L$ (or $K_i = P$), we select the subrule Ł. of $(\star L)$ (the subrule P. of $(\star L)$, respectively), and depending on the relation of $\alpha_i^{K_i}, \beta_i^{K_i}$ to these $\alpha_j^{K_j}, \beta_j^{K_j}$ that are the values of parameters occurring on $n_l$, we select the node, say $n'_l$, resulting from Cases 1-5. We then assign to the parameters at $n'_l$ that do not occur in $n_l$, say $a, b$ with $a < b$, values $\rho(a) = \alpha_i^{K_i}, \rho(b) = \beta_i^{K_i}$. Suppose now there is not $i \in \mathcal{I}$ such that $V_\star(\psi_1), V_\star(\psi_2) \in [\alpha_i^{K_i}, \beta_i^{K_i}]$. We know the relations among $V_\star(\psi_1), V_\star(\psi_2)$, and the values of parameters occurring in $n_l$, thus we know which of Cases 6-11 match these values in the model $\mathcal{A}$. Therefore, we can select the subsequent node. We have now selected the next node in the path from the root. The procedure terminates at a leaf, where there are no $L_1$-terms occurring, at which point we selected all nodes in branch $\mathcal{B}$. We also partially defined the function $\rho$. To the parameters that have not received values under $\rho$ in this procedure, we assign arbitrary values from $[0,1]$. We have now constructed an $L_0$-structure $\mathcal{M}$.

(2) To be able to show that $\mathcal{M}, \sigma \models \bigwedge l_{\mathcal{B}}$, where $l_{\mathcal{B}}$ is the leaf of branch $\mathcal{B}$, it is sufficient to prove that for all nodes in $\mathcal{B}$, $n_{\mathcal{B}}$, $\mathcal{M}_{\mathcal{B}}, \sigma \models \bigwedge n_{\mathcal{B}}$. We will sketch the proof by induction on $d(n_{\mathcal{B}}, \mathcal{B}) = h(r, \mathcal{B}) - h(n_{\mathcal{B}}, \mathcal{B})$, where $r$ is the root of tableau $\mathcal{T}$ of which $\mathcal{B}$ is a branch. We need to show that if $\mathcal{M}_{\mathcal{B}}, \sigma \models \bigwedge n_{\mathcal{B}}$,

[4] Note that $T$ may be a parameter and that this parameter has already been assigned a value by $\rho$.

then $\mathcal{M}_{\mathcal{B}}, \sigma \models \bigwedge n'_{\mathcal{B}}$, where $d(n'_{\mathcal{B}}, \mathcal{B}) = d(n_{\mathcal{B}}, \mathcal{B}) + 1$. By inspecting $n_{\mathcal{B}}, n'_{\mathcal{B}}$, we know which formula is in $n_{\mathcal{B}} - n'_{\mathcal{B}}$. Suppose that $\{z \Diamond T\} = n_{\mathcal{B}} - n'_{\mathcal{B}}$, where $\tau(\psi_1) \star \tau(\psi_2)$ is the active formula; the other cases are similar. By (1) above, we know which case and which subrule of which branch expansion rule are used to generate $n'_{\mathcal{B}}$. Thus, suppose that it was Rule ($\star$L) Ł. Case 1 (again, the other cases are similar). Let $a, b$ be the new parameters occurring at $n'_{\mathcal{B}}$. Thus, $\{0 \leq a < b \leq a_0^L\} \subseteq n'_{\mathcal{B}} - n_{\mathcal{B}}$, where $a_0^L$ is a parameter occurring at $n_{\mathcal{B}}$. By (1) above, we know that $0 \leq \rho(a) < \rho(b) \leq \rho(a_0^L)$. The other elements of $n'_{\mathcal{B}} - n_{\mathcal{B}}$ are (a) $a \leq \tau(\psi_1) \leq b$, (b) $a \leq \tau(\psi_2) \leq b$, (c) $z[\max\{a, \tau(\psi_1) + \tau(\psi_2) - b\}/\tau(\psi_1) \star \tau(\psi_2)] \Diamond T$.

*Claim.* Let $\theta$ be a subformula of $\psi$. Then $[\![\tau(\theta)]\!]^{\mathcal{M}_{\mathcal{B}}, \sigma} = V_\star(\theta)$.

*Proof of the claim.* By induction on $\theta$. The base case for atomic $\theta$ is easy. Take subformulas of $\psi$, $\theta, \varphi$ and assume the induction hypothesis for them. We show the result for $\theta \& \varphi$, i.e. $[\![\tau(\theta \& \varphi)]\!]^{\mathcal{M}_{\mathcal{B}}, \sigma} = V_\star(\theta \& \varphi)$. The cases for $\theta \to \varphi, \theta \vee \varphi, \theta \wedge \varphi$ are similar.

LHS$=[\![\tau(\theta \& \varphi)]\!]^{\mathcal{M}_{\mathcal{B}}, \sigma} = [\![\tau(\theta) \star \tau(\varphi)]\!]^{\mathcal{M}_{\mathcal{B}}, \sigma} = [\![\tau(\theta)]\!]^{\mathcal{M}_{\mathcal{B}}, \sigma} \star_{\mathcal{B}} [\![\tau(\varphi)]\!]^{\mathcal{M}_{\mathcal{B}}, \sigma} = V_\star(\theta) \star_{\mathcal{B}} V_\star(\varphi)$, by inductive hypothesis. RHS$=V_\star(\theta \& \varphi) = V_\star(\theta) \star V_\star(\varphi)$. Now, by construction in (1), $\star_{\mathcal{B}}, \star$ agree on $V_\star(\theta), V_\star(\varphi)$ as long as $\theta, \varphi$ are subformulas of $\psi$. Therefore, LHS=RHS. This completes the proof of the claim.

By (1) above and Claim, the inequalities (a), (b) and (c) are true in $\mathcal{M}_{\mathcal{B}}, \sigma$. Thus, $\mathcal{M}_{\mathcal{B}}, \sigma \models \bigwedge n'_{\mathcal{B}}$. Therefore, we showed also that $\mathcal{M}, \sigma \models \bigwedge n_{\mathcal{B}} \restriction_{L_0}$. We have now proved that in particular $\mathcal{M}, \sigma \models \bigwedge l_{\mathcal{B}}$. Thus, there is an $L_0$-structure modelling $l_{\mathcal{B}}$, $\mathcal{M}$, and a mapping $\sigma$ such that $\mathcal{M}, \sigma \models \bigwedge l_{\mathcal{B}}$. Therefore $\mathcal{B}$ is open. □

**Theorem 12.** (Soundness and Completeness)
Let $\psi$ be a formula of BL. Then every tableau with the root $\{\tau(\psi) < 1\}$ is closed iff $\psi$ is valid with respect to every continuous t-norm.

*Proof.* Immediate from lemma 11. □

## 6 Conclusion and Further Research

In this paper we constructed a tableau calculus for continuous t-norms based on the decomposition theorem and we showed that it is sound and complete. We can see that if a formula is not valid in BL, we can find a continuous t-norm that is the ordinal sum of *finitely many* proto-t-norms, in which the value of the formula is less than 1. It should be noted that the approach can be simplified due to the result of Montana in [10] who proved that the variety of BL-algebras is generated as a quasivariety by the class of all ordinal sums of finitely many Łukasiewicz t-norm. We decided to leave it in this current form as it provides a foundation for constructing tableaux for fuzzy logics whose semantics is based on ordinary sums with some components that are Product proto-t-norms. The tableau calculus may constitute a basis for further research in tableaux methods for other fuzzy logics including predicate logic or modal logic. One may also

provide the implementation for the tableau methods. Further work may involve achieving complexity results for this proof system and using it to prove the Co-NP completeness of BL.

## References

1. Annellis, I.: From semantic tableaux to Smullyan trees: the history of the falsifiability tree method. Modern Logic 1(1), 36–69 (1990)
2. Bova, S., Montagna, F.: Proof Search in Hájek's Basic Logic. ACM Transactions on Computational Logic 9(3), Article 21 (June 2008)
3. Cignoli, R., Esteva, F., Godo, L., Torrens, A.: Basic Fuzzy Logic is the logic of continuous t-norms and their residua. Soft Computing 4, 106–112 (2000)
4. Cintula, P., Hájek, P., Noguera, C.: Handbook of Mathematical Fuzzy Logic, vols. 1 & 2. College Publications (2011)
5. Fitting, M.: Introduction. In: D'Agostino, M., Gabbay, D.M., Hähnle, R., Posegga, J. (eds.) Handbook of Tableau Methods, pp. 1–44. Kluwer Academic Publishers (1999)
6. Hájek, P.: Basic fuzzy logic and BL-algebras. Soft Computing 2, 124–128 (1998)
7. Hájek, P.: Metamathematics of Fuzzy Logic. Kluwer Academic Publishers (1998)
8. Kułacka, A., Pattinson, D., Schröder L.: Syntactic Labelled Tableaux for Lukasiewicz Fuzzy ALC. In: IJCAI (2013)
9. Metcalfe, G., Olivetti, N., Gabbay, D.: Proof Theory for Fuzzy Logics. Springer (2009)
10. Montana, F.: Generating the variety of BL-algebras. Soft Computing 9, 869–874 (2005)
11. Mostert, P.S., Shields, A.L.: On the structure of semigroups on a compact manifold with boundary. Annals of Mathematics 65, 117–144 (1957)
12. Olivetti, N.: Tableaux for Łukasiewicz Infinite-valued Logic. Studia Logica 73, 81–111 (2003)
13. Orłowska, E., Golińska-Pilarek, J.: Dual Tableaux: Foundations, Methodology, Case Studies. Springer (2011)
14. Tarski, A.: A decision method for elementary algebra and geometry. University of California Press, Berkeley (1951)
15. Vetterlein, T.: Analytic Calculi for Logics of Ordinal Multiples of Standard t-Norms. Journal of Logic and Computation 18(1), 35–57 (2007)
16. Vidal, A., Bou, F., Godo, L.: An SMT-Based Solver for Continuous t-norm Based Logics. In: Hüllermeier, E., Link, S., Fober, T., Seeger, B. (eds.) SUM 2012. LNCS, vol. 7520, pp. 633–640. Springer, Heidelberg (2012)

# Possibilistic vs. Relational Semantics for Logics of Incomplete Information

Mohua Banerjee[1], Didier Dubois[2], and Lluis Godo[3]

[1] Dept. of Mathematics and Statistics, Indian Institute of Technology, Kanpur, India
[2] IRIT-CNRS, Université Paul Sabatier, Toulouse, France
[3] IIIA-CSIC, Bellaterra, Spain

**Abstract.** This paper proposes an extension of the MEL logic to a language containing modal formulae of depth 0 or 1 only. MEL is a logic of incomplete information where an agent can express both beliefs and explicitly ignored facts, that only uses modal formulae of depth 1, and no objective ones. The extended logic, called MEL$^+$ has the same axioms as, and is in some sense equivalent to, S5 with a restricted language, but with the same expressive power. The semantics is not based on Kripke models with equivalence relations, but on pairs made of an interpretation (representing the real state of facts) and a non-empty set of possible interpretations (representing an epistemic state). Soundness and completeness are established. We provide a rationale for using our approach when an agent reasons about what is known of the epistemic state of another agent and compares it with what is known about the real world. Our approach can be viewed as an alternative to the basic epistemic logic not concerned with introspection. We discuss the difference with S5 used as a logic for rough sets, and the similarity with some previous non-monotonic logics of knowledge is highlighted.

**Keywords:** Modal logic, possibility theory, epistemic logic, rough sets.

## 1 Introduction

In the recent past [1,2], some efforts have been made to relate possibility theory and modal logic under the simplest possible syntactic and semantic framework. The resulting language of this logic called MEL is a fragment of KD where no objective formula (not preceded by a modality) appears, and where modalities cannot be nested. Models of MEL are simply non-empty subsets of interpretations of some standard propositional language. They represent the possible epistemic states of some agent. Then the necessity modality represents belief, and an agent believes a proposition if and only if the latter is true in all propositional interpretations compatible with the agent's epistemic state. This logic has axioms K and D, and another axiom saying that the agent believes all tautologies. Each epistemic state can be interpreted as a Boolean possibility distribution, and it can be shown that the necessity modality is a Boolean necessity measure. This attempt to relate possibility theory and modal logic contrasts with other

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 335–344, 2014.

previous connections between these notions, involving more elaborate constructions [8,9,3,10].

The MEL logic has a language which is a fragment of the language of the logic S5 often used as an epistemic logic. Its semantics does not use accessibility relations explicitly, and is much simpler. Moreover one cannot express the introspective axioms 4 and 5 in MEL as the nesting of modalities is not possible in its language. Nevertheless, two issues are worth investigating in connection with MEL:

- Is it possible to extend the language of MEL to objective non-modal formulae, while preserving the same style of semantics as MEL ?
- Given that standard epistemic logics such as S5 rely on accessibility (equivalence) relations, what is the connection between these two semantics ?

The aim of this paper is to provide an answer to both questions. In the next section, a brief presentation of the logic MEL is recalled. Then an extension of the MEL language to objective formulae and the corresponding extension of the MEL semantics is proposed. Finally we discuss the usual Kripke style semantics of S5 and compare them to our semantics, which we consider more natural than the relational semantics based on indiscernible possible worlds according to an equivalence relation. We claim that the latter semantics is more fit to rough sets and can account for the idea of forgetting [11]. Moreover, we show that our approach comes closer to some knowledge logics proposed in the early 1990's, in the area of non-monotonic reasoning.

## 2 MEL, A Simple Epistemic Logic

The usual truth values *true* (1) and *false* (0) assigned to propositions are of ontological nature (which means that they are part of the definition of what we call *proposition*), whereas assigning to a proposition a value whose meaning is expressed by the word *unknown* sounds like having an epistemic nature: it reveals a knowledge state according to which the truth value of a proposition (in the usual Boolean sense) in a given situation is out of reach (for instance one cannot compute it, either by lack of computing power, or due to a sheer lack of information). It corresponds to an epistemic state for an agent that can neither assert the truth of a Boolean proposition nor its falsity.

Admitting that the concept of "unknown" refers to a knowledge state rather than to an ontic truth value, we may keep the logic Boolean and add to its syntax the capability of stating that we ignore the truth value (1 or 0) of propositions. The natural framework to syntactically encode statements about knowledge states of propositional logic (PL) statements is modal logic, and in particular, the logic KD. Nevertheless, if one only wants to reason about e.g. the beliefs of another agent, a very limited fragment of this language is needed. The logic MEL [1,2] was defined for that purpose.

Let us consider $\mathcal{L}$ to be a standard propositional language built up from a finite set of propositional variables $\mathcal{V} = \{p_1, \ldots, p_k\}$ along with the Boolean

connectives of conjunction and negation $\neg$. As usual, a disjunction $\varphi \vee \psi$ stands for $\neg(\neg\varphi \wedge \neg\psi)$ and an implication $\varphi \to \psi$ stands for $\neg\varphi \vee \psi$. Further we use $\top$ to denote $\varphi \vee \neg\varphi$, and $\bot$ to denote $\neg\top$. Let us consider another propositional language $\mathcal{L}_\Box$ whose set of propositional variables is of the form $\mathcal{V}_\Box = \{\Box\varphi \mid \varphi \in \mathcal{L}\}$ to which the classical connectives can be applied. It is endowed with a modality operator expressing certainty, that encapsulates formulae in $\mathcal{L}$. In other words $\mathcal{L}_\Box = \{\Box\alpha : \alpha \in \mathcal{L}\} \mid \neg\Phi \mid \Phi \wedge \Psi$.

MEL is a logic on the language $\mathcal{L}_\Box$ and with the following semantics. Let $\Omega$ be the set of classical interpretations for the propositional language $\mathcal{L}$, i.e. $\Omega$ consists of the set of mappings $w : \mathcal{L} \to \{0, 1\}$ conforming to the rules of classical propositional logic. For a propositional formula $\varphi \in \mathcal{L}$ we will denote by $Mod(\varphi)$ the set of $w \in \Omega$ such that $w(\varphi) = 1$. Models (or interpretations) for MEL correspond to epistemic states, which are simply subsets $\emptyset \neq E \subseteq \Omega$. The truth-evaluation rules of formulas of $\mathcal{L}_\Box$ in a given epistemic model $E$ is defined as follows:

- $E \models \Box\varphi$ if $E \subseteq Mod(\varphi)$
- $E \models \neg\Phi$ if $E \not\models \Phi$
- $E \models \Phi \wedge \Psi$ if $E \models \Phi$ and $E \models \Psi$

Note that contrary to what is usual in modal logic, we do not evaluate modal formulas on particular interpretations of langage $\mathcal{L}$ because modal formulas in MEL do not refer to the actual world.

The notion of logical consequence is defined as usual $\Gamma \models \Phi$ if, for every epistemic model $E$, $E \models \Phi$ whenever $E \models \Psi$ for all $\Psi \in \Gamma$.

MEL can be axiomatized in a rather simple way (see [2]). The following are a possible set of axioms for MEL in the language of $\mathcal{L}_\Box$:

(PL) Axioms of PL for $\mathcal{L}_\Box$-formulas
(K) $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
(D) $\Box\varphi \to \Diamond\varphi$
(Nec) $\Box\varphi$, for each $\varphi \in \mathcal{L}$ that is a PL tautology, i.e. if $Mod(\varphi) = \Omega$.

The only inference rule is modus ponens. The corresponding notion of proof, denoted by $\vdash$, is defined as usual from the above set of axioms and modus ponens.

This set of axioms provides a sound and complete axiomatization of MEL, that is, it holds that, for any set of MEL formulas $\Gamma \cup \{\varphi\}$, $\Gamma \models \varphi$ iff $\Gamma \vdash \varphi$. This is not surprizing: MEL is just a standard propositional logic with additional axioms, whose propositional variables are the formulas of another propositional logic, and whose interpretations are subsets of interpretations of the latter.

Notice that MEL also actually captures the $\mathcal{L}_\Box$-fragment of the normal modal logics KD, hence of other logics, such as the well-known logics KD45 and S5, also commonly referred to as the logics of belief and knowledge, respectively. However, they are obtained from KD with axioms (called 4, 5, T) that cannot be expressed in the MEL language $\mathcal{L}_\Box$.

## 3 Extending MEL to Reason about The Actual World and Someone's Beliefs

In this section we extend the language of MEL to allow dealing with not only subjective formulas that express an agent's beliefs, but also objective formulas (i.e. non-modal formulas) that express propositions that hold true in the actual world (whatever it might be). The extended language will be denoted by $\mathcal{L}^+_\Box$, and it thus contains both propositional and modal formulas. It exactly corresponds to the non-nested fragment of the language of usual modal logic.

### 3.1 Language, Axioms and Semantics

More precisely, the language $\mathcal{L}^+_\Box$ of MEL$^+$ extends $\mathcal{L}_\Box$ and is defined by the following formation rules:

- If $\varphi \in \mathcal{L}$ then $\varphi, \Box\varphi \in \mathcal{L}^+_\Box$
- If $\Phi, \Psi \in \mathcal{L}^+_\Box$ then $\neg\Phi, \Phi \wedge \Psi \in \mathcal{L}^+_\Box$

$\Diamond\varphi$ is defined as an abbreviation of $\neg\Box\neg\varphi$. Note that $\mathcal{L} \subseteq \mathcal{L}^+_\Box$ and that in $\mathcal{L}^+_\Box$ there are no formulas with nested modalities.

Semantics for MEL$^+$ are given now by "pointed" MEL epistemic models, i.e. by structures $(w, E)$, where $w \in \Omega$ and $\emptyset \neq E \subseteq \Omega$. The truth-evaluation rules of formulas of $\mathcal{L}^+_\Box$ in a given structure $(w, E)$ is defined as follows:

- $(w, E) \models \varphi$ if $w(\varphi) = 1$, in case $\varphi \in \mathcal{L}$
- $(w, E) \models \Box\varphi$ if $E \subseteq Mod(\varphi)$
- usual rules for $\neg$ and $\wedge$

Logical consequence, as usual: $\Gamma \models \Phi$ if, for every structure $(w, E)$, $(w, E) \models \Phi$ whenever $(w, E) \models \Psi$ for all $\Psi \in \Gamma$.

The following are the axioms for MEL$^+$ in the language of $\mathcal{L}^+_\Box$:

(PL) Axioms of propositional logic
(K) $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
(D) $\Box\varphi \to \Diamond\varphi$
(Nec) $\Box\varphi$, for each $\varphi \in \mathcal{L}$ that is a PL tautology, i.e. if $Mod(\varphi) = \Omega$.

The only inference rule is *modus ponens*.[1]

### 3.2 Completeness

In what follows, we will denote by $\vdash_{PL}$ the notion of proof of classical propositional language on the language $\mathcal{L}^+_\Box$ taking all $\Box$-formulas as new propositional variables. We will further let $\Gamma \cup \{\Phi\}$ be a set of $\mathcal{L}^+_\Box$-formulas. We need first to recall the following lemma [2].

[1] An equivalent presentation could be to replace (Nec) by the usual Necessitation rule in modal logics, but restricted to tautologies of propositional logic: if $\varphi \in \mathcal{L}$ is a theorem, derive $\Box\varphi$.

**Lemma 1.** *$\Gamma \vdash \Phi$ iff*
*$\Gamma \cup \{\Box\varphi \mid \vdash_{PL} \varphi\} \cup \{$instances of axioms (K), (D) and (Nec)$\} \vdash_{PL} \Phi$*

**Theorem 1.** *$\Gamma \vdash \Phi$ iff $\Gamma \models \Phi$.*

*Proof.* From left to right is easy, as usual. For the converse direction, assume $\Gamma \nvdash \Phi$. By the preceding lemma and the completeness of PL, there exists a propositional evaluation $v$ on the whole language $\mathcal{L}^+_\Box$ (taking $\Box$-formulas as genuine propositional variables) such that $v(\Psi) = 1$ for all $\Psi \in \Gamma \cup \{\Box\varphi \mid \vdash_{PL} \varphi\} \cup \{$instances of axioms (K) and (D)$\}$ but $v(\Phi) = 0$. We have to build a structure $(w, E)$ that it is a model of $\Gamma$ but not of $\Phi$. So, we take $(w, E)$ as follows:

- $w$ is defined as the restriction of $v$ to $\mathcal{L}$, i.e. $w(\varphi) = v(\varphi)$ for all $\varphi \in \mathcal{L}$.
- $E = \bigcap\{Mod(\varphi) \mid v(\Box\varphi) = 1\}$

Note that, since by assumption $v(\Diamond\top) = 1$, $E \neq \emptyset$. Then the last step is to show that, for every $\Psi \in \mathcal{L}^+_\Box$, $v(\Psi) = 1$ iff $(w, E) \models \Psi$. We prove this by induction. The case $\Psi$ being a non-modal formula from $\mathcal{L}$ is clear, since in that case $w(\Psi) = v(\Psi)$. The interesting case is when $\Psi = \Box\psi$. Then we have:

(i) If $v(\Box\psi) = 1$ then, by definition of $E$, $E \subseteq Mod(\psi)$, and hence $(w, E) \models \Box\psi$.
(ii) Conversely, if $E \subseteq Mod(\psi)$, then there must exist $\gamma$ such that $v(\Box\gamma) = 1$ and $Mod(\gamma) \subseteq Mod(\psi)$. Hence this means that $\gamma \to \psi$ is a PL theorem, and hence we have first, by the necessitation axiom, that $v(\Box(\gamma \to \psi)) = 1$, and thus $v(\Box\gamma) \leq v(\Box\psi)$ holds as well by axiom (K), and therefore $v(\Box\psi) = 1$ holds as well.

As a consequence, we have that $(w, E) \models \Psi$ for all $\Psi \in \Gamma$ and $(w, E) \not\models \Phi$.

*Remark 1.* Notice that if the notion of logical consequence $\models$ is reduced to considering only structures $(w, E)$ such that $w \in E$, then one should add the following well-known axiom (T): $\Box\varphi \to \varphi$ to keep completeness.

## 4 Relating MEL and MEL$^+$ to KD45 and S5

Recall the normal modal systems KD, KD4, KD45 and S5 (see e.g. [4] for details).

**Proposition 1.** *Let $\varphi$ a formula from $\mathcal{L}_\Box$. Then MEL $\vdash \varphi$ iff KD $\vdash \varphi$.*

*Proof.* Assume $KD \nvdash \varphi$, then there is a serial Kripke model $(W, e, R)$ and $w \in W$ such that $e(w, \varphi) = 0$. Since $\varphi$ does not contain nested modal operators, $e(w, \varphi)$ only depends on the truth-evaluations of subformulas of $\varphi$ at all the worlds from $R(w) = \{w' \mid wRw'\}$, which is non-empty since $R$ is serial. We can assume $R(w) \subseteq \Omega$. Define, for each $w' \in R(w)$, the propositional evaluation $v_{w'}(.) := e(w', .)$, and the epistemic model $E_w := \{v_{w'} : w' \in R(w)\}$. Then $E_w \models \varphi$, if and only if $e(w, \varphi) = 1$. Hence MEL $\nvdash \varphi$.

Conversely, assume MEL $\nvdash \varphi$, then there is an epistemic model $E$ such that $E \not\models \varphi$. Now, consider the Kripke model $M = (E, e, R)$ where $R = E \times E$ with $e(v, \cdot) = v(\cdot)$ for every $v \in E$. $M$ is clearly a serial model. The fact that $E \not\models \varphi$ implies that $e(w, \varphi) = 0$ for some $w \in E$, and hence $M \not\models \varphi$, and by completeness, $KD \nvdash \varphi$.

**Corollary 1.** *Let $\varphi$ a formula from $\mathcal{L}_\Box$. Then MEL $\vdash \varphi$ iff $L \vdash \varphi$ for $L \in \{KD4, KD45, S5\}$.*

Now let us consider the case of MEL$^+$. Then we have even stronger relationships to KD45 and S5.

**Proposition 2.** *Let $\varphi$ a formula from $\mathcal{L}^+_\Box$. Then MEL$^+ \vdash \varphi$ iff $KD \vdash \varphi$.*

*Proof.* The proof is very similar to that of Proposition 1.

**Corollary 2.** *Let $\varphi$ a formula from $\mathcal{L}^+_\Box$. Then MEL$^+ \vdash \varphi$ iff $L \vdash \varphi$ for $L \in \{KD4, KD45\}$.*

Note that this corollary does not hold for L = S5, indeed, $\Box\varphi \to \varphi$ is an axiom of S5 that is not provable in MEL$^+$.

Let us call MEL$^{++}$ the extension of MEL$^+$ with the axiom (T): $\Box\varphi \to \varphi$. Then notice that an easy adaptation of the proof of completeness theorem for MEL$^+$ proves that MEL$^{++}$ is complete with respect to the class of *reflexive* pointed epistemic models $(w, E)$ where $w \in E$.

**Proposition 3.** *Let $\varphi$ a formula from $\mathcal{L}^+_\Box$. Then, MEL$^{++} \vdash \varphi$ iff $S5 \vdash \varphi$.*

The proof easily follows from that of Proposition 2 by taking into account that one has to deal with reflexive epistemic models.

Moreover, by recalling the well-known result that any formula of KD45 and S5 is logically equivalent to another formula without nested modalities, we can formulate the following stronger relationships.

**Proposition 4.** *The following conditions hold true:*

- *For any arbitrary modal formula $\varphi$, there is a formula $\varphi' \in \mathcal{L}^+_\Box$ such that $KD45 \vdash \varphi$ iff MEL$^+ \vdash \varphi'$.*
- *For any arbitrary modal formula $\varphi$, there is a formula $\varphi' \in \mathcal{L}^+_\Box$ such that $S5 \vdash \varphi$ iff MEL$^{++} \vdash \varphi'$.*

*Remark 2.* Recently, Petruszczak [16] indicated that simplified Kripke frames could indeed be used for the semantics of systems K45, KB5 and KD45, using subsets of propositional valuations in place of relations, as we proposed. He proves it by constructing specific accessibility relations equivalent to such subsets, as in [1] for MEL, while the completeness proof in [2] and here is direct.

## 5 MEL$^+$ vs. Other Logics of Incomplete Information

The language MEL is supposed to encode the following situation [1]. There are two agents, one of which, say $\mathcal{A}$, reasons about some beliefs possessed by another agent $\mathcal{B}$, the former is aware of, on the basis of the testimony of the latter. Namely, $\mathcal{A}$ partially knows what the other agent believes. A belief base in MEL typically contains the testimony of agent $\mathcal{B}$, namely propositions agent

$\mathcal{B}$ believes ($\Box\alpha$), some that he explicitly does not know ($\Diamond\alpha \wedge \Diamond\neg\alpha$), and finally some other propositions that agent $\mathcal{A}$ is aware the agent $\mathcal{B}$ knows the truth-value of, without guessing which ($\Box\alpha \vee \Box\neg\alpha$). The logic MEL enables agent $\mathcal{A}$ to infer more beliefs agent $\mathcal{B}$ possesses but did not reveal. Such a (meta-)belief base for agent $\mathcal{A}$ is equivalent to a set of possible epistemic states for agent $\mathcal{B}$. In MEL$^+$, agent $\mathcal{A}$ is allowed to add what is known about the real world in the form of standard propositions. So $\alpha \wedge \Box\neg\alpha$ means that agent $\mathcal{A}$ considers $\alpha$ is true, while he knows that agent $\mathcal{B}$ believes it is false. Under this set-up, a MEL$^+$ model $(w, E)$ is interpreted as the fact that $\mathcal{A}$ envisages the real world to be $w$ and the epistemic state of $\mathcal{B}$ to be $E$. If $\mathcal{A}$ considers that $\mathcal{B}$'s beliefs are always correct, the former can assume axiom T is valid, thus he reasons in MEL$^{++}$ to strengthen his own knowledge of the real world. Alternatively, $\mathcal{A}$ may mistrust $\mathcal{B}$ and may wish to take advantage of knowing wrong beliefs of $\mathcal{A}$.

### 5.1 Epistemic Logic and Accessibility Relations: A Critique

In contrast, usual semantics of S5 [7] consider the epistemic state of an agent is modelled by an equivalence relation $R$ on a set of possible worlds $W$.[2] The statement $wRw'$ reads "world $w'$ is accessible from $w$". The world $w'$ is said to be an epistemic or doxastic alternative to world $w$ for the agent, depending on whether knowledge or belief is the considered attitude. There are various attempts to make sense of this relation, such that the agent cannot distinguish $w$ from $w'$, or $w'$ is a possible state of affairs from the point of view of what the agent knows in $w$, etc. The underlying idea seems to be that "the set of worlds considered possible by an agent depends on his or her informational resources at that instant" (Stanford Encyc. Philos.).

However this view, which seems to be shared by many scholars, is not so easy to grasp. Interpreting "accessible worlds" as worlds compatible with the agent epistemic state, we can assume that the epistemic state of the agent depends on his or her informational resources at that instant (in MEL this possibility is not considered). But it is not clear that the agent is aware of his own informational resources to the point of articulating them in the same language as the one he uses to speak about the current states of affairs. If $w$ stands for an objective state of facts, it may not include the particulars of the agent. The epistemic state of the agent depends on many hidden internal features of the agent but his current observations about the actual world $w$ refer to something external, the agent is focused on. It is not clear that the vocabulary used to describe the actual world $w$ is rich enough to also account for the inner state (of health or informational resources) of the agent that holds some beliefs about $w$. For instance suppose the agent has incomplete information on the outcome of coin flipping round after the toss: this epistemic state will generally not depend on the outcome of the toss (like "if the result is heads then he knows it, otherwise he does not know").

[2] See [6] for an alternative semantics that makes the internal structure of possible worlds more explicit, encoding both objective facts and agents' mental states in a possible world. We thank one reviewer for pointing this out to us.

In a nutshell, while an equivalence class of $R$ represents context-dependent knowledge of the agent, it is not clear that this contextual dependence is part of the agent's knowledge about himself, let alone about another agent. So in the epistemic logic approach the accessibility relation seems to be a circular notion, where possible worlds seem to include the description of the agent mental circumstances as well as the description of his epistemic state regarding the problem he considers. This view may make sense when introspection is the main issue (the agent being partially unaware of his own knowledge), but it seems to be at odds with the problem of an agent reflecting about other agents knowledge, as in the set-up for MEL and MEL$^+$. As the set-up for MEL is not introspective, this relational semantics looks like a questionable artifact for this logic, where we assume agents are aware of their own knowledge and lack of knowledge. Note that if $R = W \times W$ one can only distinguish between tautologies (i.e. $\Box\phi$ where $\phi$ is a PL tautology), contradictions and contingent modal propositions.

One may extend the MEL set-up by considering a separate set of possible *mental dispositions* $S$ corresponding to "informational resources" (due to specific situations or circumstances) an agent can access at a particular moment. On the other hand, $W$ encodes the question the agent is concerned with at that moment; it pertains to the outside world, so $S \cap W = \emptyset$ The accessibility relation $R$ is relating $S$ to $W$, namely $E = R(s) \subseteq W$ is the epistemic state of the agent when his mental disposition is $s$. Note that, under this view, there is no point of $R$ being an equivalence relation. But this extension assumes that the set $S$ of mental dispositions of the agent is known and observable by another agent.

### 5.2 Comparison to Logics of Rough Sets

The semantics of S5 in terms of equivalence relations [4] makes it in fact the natural logical setting for rough sets [14]. Pawlak's rough sets [15] are based on the notion of *approximation spaces* $(W, R)$, where $R$, called the *indiscernibility* relation, is an equivalence relation on the domain $W$ of discourse. The premise is that due to lack of complete information about the objects in the domain, it is likely that many of the objects are indistinguishable from each other. This is patent in *information systems* $I := (W, At, V_{At}, f)$, where $At$ is a set of attributes, $V_{At}$ a set of values for the attributes in $At$, and $f : W \times At \to V_{At}$ a function assigning values for attributes to objects of the domain. $I$ then induces an indiscernibility relation $R$ corresponding to every subset $B \subseteq At$:

$$xRy, \text{ if and only if } f(x, b) = f(y, b), \text{ for all } b \in B.$$

The lower and upper approximations of a subset $X$ of the domain with respect to $R$ are defined as

$$\underline{X} := \{x \in W : R(x) \subseteq X\};\ \overline{X} := \{x \in W : R(x) \cap X \neq \emptyset\}.$$

Sets with identical approximations are said to be *roughly equal*, and for any $X$, the collection of all subsets of the domain roughly equal to $X$ is termed a rough set by Pawlak in [15].

Any logic of rough sets thus has an essential modal nature, the necessity and possibility operators capturing the lower and upper approximations in the rough set semantics respectively. In fact, for any S5 Kripke model $M = (W, e, R)$, one may observe that $e(\Box\varphi) = \underline{e(\varphi)}$ and $e(\Diamond\varphi) = \overline{e(\varphi)}$, where $e(\varphi) := \{w \in W : e(w, \varphi) = 1\}$. However, unlike MEL or MEL$^+$, rough set logics make use of the full modal language, that is, nested modalities are allowed. For instance one would use a formula such as $\Diamond(p \wedge \Box q)$ to refer to a set $\overline{(X \cap \underline{Y})}$.

This approach can easily be extended to rough set models based on a relation that is not necessarily an equivalence one [17,18]. These logics remain modal, and use nested modalities. Indeed, it is well known in modal logic [4] that, once fixed the basic axioms (PL) and (K), then each additional modal axioms corresponds a different property of the accessibility relation.

### 5.3 Comparison to the Logic of Minimal Belief and Negation as Failure (MBNF)

In [12] Lifschitz defines a simplified version of Lin and Shoham logic [13] of minimal knowledge and justified assumptions. His (nonmonotonic) logic, in the propositional version, contains two modal operators, one for minimal belief $B$ and another for negation as failure *not*. For *positive* formulas, i.e. formulas not containing *not*, the (monotonic) semantics is very similar to that of MEL$^+$: semantics are given by structures $(I, S)$, where $I$ is an interpretation of propositional variables (or equivalently a set of atoms) and $S$ a set of interpretations. The author writes that,"intuitively, $I$ represents 'the real world', and $S$ the set of 'possible worlds' ". As in MEL$^+$, for a nonmodal formula $\varphi$, a structure $(I, S)$ satisfies the formula $B\varphi$ whenever each $I' \in S$ satisfies $\varphi$. Therefore MBNF structures $(I, S)$ exactly correspond to pointed epistemic MEL$^+$ models.

The nonmonotonic semantics of MBNF is defined as to capture the notion of minimal belief. A *model* of a theory $T$ (set of positive formulas) is a structure $(I, S)$ such that it makes true all the formulas of $T$ and it is maximal in the sense that there is no other structure $(I', S')$ making the formulas of $T$ true and such that $S \subsetneq S'$. For instance, the only model of $B\varphi$ in this semantics is $(I, Mod(\varphi))$, while the models of $B\varphi \vee B\psi$ are $(I, Mod(\varphi))$ and $(I, Mod(\psi))$. Then the corresponding notion of (nonmonotonic) consequence relation is defined accordingly. For instance one has $Bp \models \neg Bq$ but $\{Bp, Bq\} \not\models \neg Bq$.

Actually, MBNF models for formulas of the language of MEL (resp. MEL$^+$) correspond to the *minimum specific* epistemic models in MEL (reps. pointed epistemic models in MEL$^+$) in the sense of possibilistic logic.

## 6 Conclusion

In this paper, we argue that the usual semantics of epistemic logics based on accessibility relations is not very natural when the purpose is to reason about beliefs or incomplete knowledge entertained by an external agent, and introspection is ruled out. To this end, we have shown that the fragment MEL$^+$ (resp.

MEL$^{++}$) of the KD45 (resp. S5) logic, the richest of doxastic (resp. epistemic) logics, involving modal formulas of depth 0 or 1 can have simplified semantics that are more intuitive than equivalence relations, while the latter make sense for capturing rough sets. The connection between MEL$^{+}$ and MBNF clearly suggests the former has more to do with logic programming than to the mainstream modal logic tradition, as already noticed with generalized possibilistic logic (a multimodal extension of MEL), that can encode answer-set programming [5].

**Acknowledgments.** Godo acknowledges partial support by the Spanish MINECO project EdeTRI TIN2012-39348-C02-01.

## References

1. Banerjee, M., Dubois, D.: A Simple Modal Logic for Reasoning about Revealed Beliefs. In: Sossai, C., Chemello, G. (eds.) ECSQARU 2009. LNCS, vol. 5590, pp. 805–816. Springer, Heidelberg (2009)
2. Banerjee, M., Dubois, D.: A simple logic for reasoning about incomplete knowledge. Int. J. Approximate Reasoning 55(2), 639–653 (2014)
3. Boutilier, C.: Modal logics for qualitative possibility theory. Int. J. Approximate Reasoning 10, 173–201 (1994)
4. Chellas, B.F.: Modal Logic: An Introduction. Cambridge University Press (1980)
5. Dubois, D., Prade, H., Schockaert, S.: Stable Models in Generalized Possibilistic Logic. In: Brewka, G., et al. (eds.) Proc. of KR 2012, pp. 519–529. AAAI Press (2012)
6. Fagin, R., Vardi, M.Y.: An internal semantics for modal logic. In: Proc. of STOC 1985, pp. 305–315. ACM, New York (1985)
7. Halpern, J.Y., Fagin, R., Moses, Y., Vardi, M.Y.: Reasoning About Knowledge. MIT Press (2003) (Revised paperback edition)
8. Fariñas del Cerro, L., Herzig, A.: A modal analysis of possibility theory. In: Jorrand, P., Kelemen, J. (eds.) FAIR 1991. LNCS, vol. 535, pp. 11–18. Springer, Heidelberg (1991)
9. Hájek, P.: A qualitative fuzzy possibilistic logic. Int. J. of Approximate Reasoning 12, 1–19 (1994)
10. Hájek, P., Harmancová, D., Esteva, F., Garcia, P., Godo, L.: On modal logics for qualitative possibility in a fuzzy setting. In: Lopez de Mántaras, R., Poole, D. (eds.) Proc. of UAI-94, pp. 278–285. Morgan Kaufmann, San Francisco (1994)
11. Lang, J., Marquis, P.: A forgetting-based approach. Artificial Intelligence 174, 799–823 (2010)
12. Lifschitz, V.: Minimal belief and negation as failure. Artificial Intelligence 70, 53–72 (1994)
13. Lin, F., Shoham, Y.: A logic of knowledge and justified assumptions. Artificial Intelligence 57, 271–289 (1992)
14. Orlowska, E.: A logic of indiscernibility relations. In: Skowron, A. (ed.) SCT 1984. LNCS, vol. 208, pp. 177–186. Springer, Heidelberg (1985)
15. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11(5), 341–356 (1982)
16. Petruszczak, A.: Simplified Kripke style semantics for modal logics K5, KB5 and KD45. Bulletin of the Section of Logic 38, 163–171 (2009)
17. Yao, Y., Lin, T.: Generalization of rough sets using modal logic. Automation and Soft Computing 2(2), 103–120 (1996)
18. Yao, Y., Wang, S., Lin, T.: A review of rough set models. In: Polkowski, L., Skowron, A. (eds.) Rough Sets and Data Mining: Analysis for Imprecise Data, pp. 47–75. Kluwer Academic Publishers (1997)

# Resolution in Linguistic First Order Logic Based on Linear Symmetrical Hedge Algebra

Thi-Minh-Tam Nguyen[1], Viet-Trung Vu[2],
The-Vinh Doan[2], and Duc-Khanh Tran[3]

[1] Vinh University, Vietnam
nmtam@vinhuni.edu.vn
[2] Hanoi University of Science and Technology, Vietnam
{trungvv91,doanthevinh1991}@gmail.com
[3] Vietnamese German University, Vietnam
khanh.td@vgu.edu.vn

**Abstract.** This paper focuses on resolution in linguistic first order logic with truth value taken from linear symmetrical hedge algebra. We build the basic components of linguistic first order logic, including syntax and semantics. We present a resolution principle for our logic to resolve on two clauses having converse linguistic truth values. Since linguistic information is uncertain, inference in our linguistic logic is approximate. Therefore, we introduce the concept of reliability in order to capture the natural approximation of the resolution inference rule.

**Keywords:** Linear Symmetrical Hedge Algebra, Linguistic Truth Value, Linguistic First Order Logic, Resolution, Automated Reasoning.

## 1 Introduction

Automated reasoning theory based on resolution rule of Robinson [14] has been research extensively in order to find efficient proof systems [1, 6]. However, it is difficult to design intelligent systems based on traditional logic while most of the information we have about the real world is uncertain. Along with the development of fuzzy logic, non-classical logics became formal tools in computer science and artificial intelligence. Since then, resolution based on non-classical logic (especially multi-valued logic and fuzzy logic) has drawn the attention of many researchers.

In 1965, Zadeh introduced fuzzy set theory known as an extension of set theory and applied widely in fuzzy logic [21]. Many researchers have presented works about the fuzzy resolution in fuzzy logic [2, 8, 9, 15, 18, 20]. In 1990, Ho and Wechler proposed an approach to linguistic logic based on the structure of natural language [10]. The authors introduced a new algebraic structure, called hedge algebra, to model linguistic truth value domain, which applied directly to semantics value in inference. There also have been many works about inference on linguistic truth value domain based on extended structures of hedge algebra

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 345–354, 2014.

such as linear hedge algebra, monotony linear hedge algebra [7, 12, 13]. Researchers also presented truth functions of new unary connectives (hedges) from the set of truth values to handle fuzzy truth values in a natural way [3, 5, 19]. Recently, we have presented the resolution procedure in linguistic propositional logic with truth value domain taken from linear symmetrical hedge algebra [11]. We have constructed a linguistic logic system, in which each sentence in terms of *"It is very true that Mary studies very well"* is presented by $P^{\mathsf{VeryTrue}}$, where P is *"Mary studies very well"*. Two clauses having converse linguistic truth values, such as $P^{\mathsf{VeryTrue}}$ and $P^{\mathsf{MoreFalse}}$, are resolved by a resolution rule. However, we cannot intervene in the structure of a proposition. For example with the knowledge base: *"It is true that if a student studies hard then he will get the good marks"* and *"It is very true that Peter studies hard"*, we cannot infer to find the truth value of the sentence *"Peter will get the good marks"*. Linguistic first order logic overcomes this drawback of linguistic propositional logic. Furthermore, knowledge in the linguistic form maybe compared in some contexts, such as when we tell about the value of linguistic variable *Truth*, we have $\mathsf{LessTrue} < \mathsf{VeryTrue}$ or $\mathsf{MoreFalse} < \mathsf{LessFalse}$. Therefore, linear symmetrical hedge algebra is an appropriate to model linguistic truth value domain.

As a continuation of our research works on resolution in linguistic propositional logic systems [11, 17], we study resolution in linguistic first order logic. We construct the syntax and semantics of linguistic first order logic with truth value domain taken from linear symmetrical hedge algebra. We also propose a resolution rule and a resolution procedure for our linguistic logic. Due to the uncertainty of linguistic information, each logical clause would be associated with a certain confidence value, called reliability. Therefore, inference in our logic is approximate. We shall build an inference procedure based on resolution rule with a reliability $\alpha$ which ensures that the reliabilities of conclusions are less than or equal to reliabilities of premises.

The paper is structured as follows: section 2 introduces basic notions of linear symmetrical hedge algebras and logical connectives. Section 3 describes the syntax and semantics of our linguistic first order logic with truth value domain based on linear symmetrical hedge algebra. Section 4 proposes a resolution rule and a resolution procedure. Section 5 concludes and draws possible future work. Due to the lack of space several proofs had to be omitted. They appear in the full paper at `http://arxiv.org/abs/1403.6237`.

## 2 Linear Symmetrical Hedge Algebra

We consider an abstract algebra called hedge algebra $HA = (X, G, H, \leq)$ where $G$ is the set of generators and $H$ is the set of hedges. The set of values $X$ generated from $G$ and $H$ is defined as $X = \{\delta c | c \in G, \delta \in H\}$. $\leq$ is a partial order on X such that $a \leq b$ if $a < b$ or $a = b$ $(a, b \in X)$.

Let $h, k$ be two hedges in the set of hedges $H$. Then $k$ is said to be *positive (negative)* w.r.t. $h$ if for every $x \in X$, $hx \geq x$ implies $khx \geq hx (khx \leq hx)$ or, conversely, $hx \leq x$ implies $khx \leq hx (khx \geq hx)$. $h$ and $k$ are *converse*

if $\forall x \in X, hx \leq x$ iff $kx \geq x$. $h$ and $k$ are *compatible* if $\forall x \in X, x \leq hx$ iff $x \leq kx$. $h$ modifies terms stronger or equal than $k$, denoted by $h \geq k$, if $\forall x \in X, (hx \geq kx \geq x)$ or $(hx \geq kx \geq x)$.

The set of primary terms $G$ usually consists of two comparable ones, denoted by $c^- < c^+$. Such HAs are called *symmetric* ones. For symmetric HAs, the set of hedges $H$ is decomposed into two non-empty disjoint subsets $H^+$ and $H^-$ so that each element in $H^+$ is a converse operation of the operations in $H^-$, i.e. $H^+ = \{h \in H | hc^+ > c^+\}$ and $H^- = \{h \in H | hc^+ < c^+\}$. Two hedges in each of the sets $H^+$ and $H^-$ are comparable. Thus, $H^+$ and $H^-$ become posets. Let $I \notin H$ be the identity hedge, i.e., $Ix = x \ \forall x \in X$. A linear symmetric HA is defined as follows:

**Definition 1.** *[10, 13] An abstract algebra $(X, G, H, \leq)$ where $G = \{c^-, c^+\}$, $H \neq \emptyset$ and $X = \{\sigma c | c \in G, \sigma \in H^*\}$, is called a linear symmetric HA (lin-HA, for short) if it is satisfies the following conditions:*

*(A1) For all $h \in H^+$ and $k \in H^i$, $h$ and $k$ are converse,*
*(A2) The sets $H^+ \cup I$ and $H^- \cup I$ are linearly ordered with the least element $I$,*
*(A3) For each pair $h, k \in H$, either $h$ is positive or negative w.r.t. $k$,*
*(A4) If $h \neq k$ and $hx \leq kx$ then $h'hx \leq k'kx$ for all $h, k, h', k' \in H$ and $x \in X$,*
*(A5) If $u \notin H(v)$ and $u \leq v$ then $u \leq hv$, for any $h \in H$.*

Let $x = h_n ... h_1 a$ be an element of the hedge algebra $AX$ where $a \in \{c^+, c^-\}$. The converse element of $x$ is an element $\overline{x}$ such that $\overline{x} = h_n ... h_1 a'$ where $a' \in \{c^+, c^-\}$ and $a' \neq a$. In *lin-HA*, every element $x \in X$ has an unique converse element in $X$.

HAs are extended by augmenting two hedges $\Phi$ and $\Sigma$ defined as $\Phi(x) = infimum(H(X))$ and $\Sigma(x) = supremum(H(x))$, for all $x \in X$ [4]. It is shown that, for a free *lin-HA* with $H \neq \emptyset, \Phi(c^+) = \Sigma(c^-)$. We denote $\Sigma(c^+) = \top$ and $\Phi(c^-) = \bot$. Let us put $W = \Phi(c^+) = \Sigma(c^-)$ (called the middle truth value), we have $\bot < c^- < W < c^+ < \top$.

**Definition 2.** *A linguistic truth domain $\overline{X}$ taken from a lin-HA $AX = (X, \{c^-, c^+\}, H, \leq)$ is defined as $\overline{X} = X \cup \{\bot, W, \top\}$, where $\bot, W, \top$ are the least, the neutral, and the greatest elements of $\overline{X}$, respectively.*

**Proposition 1.** *[4] For any lin-HA $AX = (X, G, H, \leq)$, the linguistic truth domain $\overline{X}$ is linearly ordered.*

To define the logical connectives, different t-norms and t-conorms are used [11, 16, 20]. In this paper we define logical connectives using Gödel's t-norm and t-conorm operators. Let $K = \{n | n \in \mathbb{N}, n \leq N_0\}$. Let $m, n \in K$. Gödel's t-norm (T) and t-conorm (S) operators are defined as follows:
$-T(m, n) = \min(m, n)$.
$-S(m, n) = \max(m, n)$.

Given a *lin-HA* AX, since all the values in AX are linearly ordered, truth functions for conjunctions and disjunctions are Gödel's t-norms and t-conorms, respectively.

**Definition 3.** *Let $\overline{X}$ be a linguistic truth domain, which is a lin-HA. The logical connectives $\wedge$ (respectively $\vee$) over $\overline{X}$ are defined to be Gödel's t-norm (respectively t-conorm), and furthermore to satisfy the following: $\neg\alpha = \overline{\alpha}$, and $\alpha \to \beta = (\neg\alpha) \vee \beta$, where $\alpha, \beta \in \overline{X}$.*

**Proposition 2.** *Let $\overline{X} = X \cup \{\bot, W, \top\}$ be a linguistic truth domain taken from a lin-HA $AX = (X, \{c^-, c^+\}, H, \leq)$; $\alpha, \beta, \gamma \in X$, we have:*

- *Double negation: $\neg(\neg\alpha) = \alpha$*
- *Commutative: $\alpha \wedge \beta = \beta \wedge \alpha$, $\alpha \vee \beta = \beta \vee \alpha$*
- *Associative: $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$, $(\alpha \vee \beta) \vee \gamma = \alpha \vee (\beta \vee \gamma)$*
- *Distributive: $\alpha \wedge (\beta \vee \gamma) = (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$, $\alpha \vee (\beta \wedge \gamma) = (\alpha \vee \beta) \wedge (\alpha \vee \gamma)$*

## 3 Linguistic First Order Logic Based on Linear Symmetrical Hedge Algebra

In this section we define the syntax and semantics of our linguistic first-order logic.

**Definition 4.** *The alphabet of a linguistic first-order language consists of the following sets of symbols:*

- *constant symbols: a set of symbols $a, b, c, \ldots$, each of 0-ary;*
- *logical constant symbols:* $\mathsf{MoreTrue}, \mathsf{VeryFalse}, \bot, \top, \ldots$*;*
- *variable: $x, y, z, \ldots$;*
- *predicate symbols: a set of symbols $P, Q, R, \ldots$, each associated with a positive integer n, arity. A predicate with arity n is called n-ary;*
- *function symbols: a set of symbols $f, g, h, \ldots$, each associated with a positive integer n, arity. A function with arity n is called n-ary;*
- *logical connectives: $\vee, \wedge, \neg, \to, \leftrightarrow$;*
- *quantifies: universal quantification $\forall$, existentional quantification $\exists$;*
- *auxiliary symbols: $\square, (, ), \ldots$.*

**Definition 5.** *A term is defined recursively as follows:*

- *either every constant symbol or every variable is a term,*
- *if $t_1, \ldots, t_n$ are terms and $f$ is a n-ary function symbol, $f(t_1, \ldots, t_n)$ is a term (functional term).*

**Definition 6.** *An atom is either a zero-ary predicate symbol or a n-ary predicate symbol $P(t_1, \ldots, t_n)$, where $t_1, \ldots, t_n$ are terms.*

**Definition 7.** *Let $A$ be an atom and $\alpha$ be a logical constant. Then $A^\alpha$ is called a literal.*

**Definition 8.** *Formulae are defined recursively as follows:*

- *a literal is a formula,*
- *if $F, G$ are formulae, then $F \vee G$, $F \wedge G$, $F \to G, F \leftrightarrow G, \neg F$ are formulae,*
- *if $F$ is a formula and $x$ is a free variable in $F$, then $(\forall x)F$ and $(\exists x)F$ are formulae.*

The notions of free variable, bound variable, substitution, unifier, most general unifier, ground formula, closed formula, etc. are similar to those of classical logic.

**Definition 9.** *A clause is a finite disjunction of literals represented by* $L_1 \vee L_2 \vee ... \vee L_n$, *where* $L_i (i = 1, 2, ..., n)$ *is a literal. An empty clause is denoted by* $\Box$.

A formula is in conjunctive normal form (CNF) if it is a conjunction of clauses. It is well known that transforming a formula in first order logic into a CNF formula preserves satisfiability [1]. In Section 4 we shall be working with a resolution procedure which processes CNF formulae, or equivalently clause sets.

**Definition 10.** *An interpretation for the linguistic first order logic is a pair* $I=<D,A>$ *where* $D$ *is a non empty set called domain of* $I$, *and* $A$ *is a function that maps:*

- *every constant symbol* $c$ *into an element* $c^A \in D$;
- *every n-ary function symbol* $f$ *into a function* $f^A : D^n \to X$;
- *every logical constant symbol* $l$ *into an element* $l^A \in X$;
- *every n-ary predicate symbol* $P$ *into an n-ary relation* $P^A : D^n \to X$, *where* $X$ *is the truth value domain taken from lin-HA;*
- *every variable* $x$ *into a term.*

Given an interpretation $I=<D,A>$ for the linguistic first order logic, the truth value of a symbol $S$ in the alphabet of the logic is denoted by $I(S)$.

**Definition 11.** *Given an interpretation* $I=<D,A>$, *we define:*

- *Value of a term:* $I(t) = t^A$, $I(f(t_1, \ldots, t_n)) = f(I(t_1), \ldots, I(t_n))$.
- *Truth value of an atom:* $I(P(t_1, \ldots, t_n)) = P(I(t_1), \ldots, I(t_n))$.
- *Truth value of a logical constant:* $I(c) = c^A$.
- *Let* $P$ *be an atom such that* $I(P) = \alpha_1$. *Truth value of a literal* $P^{\alpha_2}$:

$$I(P^{\alpha_2}) = \begin{cases} \alpha_1 \wedge \alpha_2 \text{ if } \alpha_1, \alpha_2 > \mathsf{W}, \\ \neg(\alpha_1 \vee \alpha_2) \text{ if } \alpha_1, \alpha_2 \leq \mathsf{W}, \\ (\neg\alpha_1) \vee \alpha_2, \text{ if } \alpha_1 > \mathsf{W}, \alpha_2 \leq W, \\ \alpha_1 \vee (\neg\alpha_2), \text{ if } \alpha_1 \leq \mathsf{W}, \alpha_2 > \mathsf{W}. \end{cases}$$

- *Let* $F$ *and* $G$ *be formulae. Truth value of a formula:*

  - $I(\neg F) = \neg I(F)$
  - $I(F \wedge G) = I(F) \wedge I(G)$
  - $I(F \vee G) = I(F) \vee I(G)$
  - $I(F \to G) = I(F) \to I(G)$
  - $I(F \leftrightarrow G) = I(F) \leftrightarrow I(G)$
  - $I((\forall x)F) = min_{\forall d \in D}\{I(F)\}$
  - $I((\exists x)F) = max_{\exists d \in D}\{I(F)\}$

**Definition 12.** *Let* $I=<D,A>$ *be an interpretation and* $F$ *be a formula. Then*

- $F$ *is true iff* $I(F) \geq W$. $F$ *is satisfiable iff there exists an interpretation* $I$ *such that* $F$ *is true in* $I$ *and we say that* $I$ *is a model of* $F$ *(write* $I \models F$*) or* $I$ *satisfies* $F$.

- *$F$ is false iff $I(F) < W$ and we say that $I$ falsifies $F$. $F$ is unsatisfiable iff there exists no interpretation that satisfies $F$.*
- *$F$ is valid iff every interpretation of $F$ satisfies $F$.*
- *A formula $G$ is a logical consequence of formulas $\{F_1, F_2, \ldots, F_n\}$ iff for every interpretation $I$, if $I \models F_1 \wedge F_2 \wedge \ldots \wedge F_n$ we have that $I \models G$.*

**Definition 13.** *Two formulae $F$ and $G$ are logically equivalent iff $F \models G$ and $G \models F$ and we write $F \equiv G$.*

It is infeasible to consider all possible interpretations over all domains in order to prove the unsatisfiability of a clause set $S$. Instead, we could fix on one special domain such that $S$ is unsatisfiable iff $S$ is false under all the interpretations over this domain. Such a domain, which is called the Herbrand universe of $S$, defined as follows.

Let $H_0$ be the set of all constants appearing in $S$. If no constant appears in $S$, then $H_0$ is to consist of a single constant, say $H_0 = \{a\}$. For $i = 0, 1, 2, \ldots$, let $H_{i+1}$ be the union of $H_i$ and the set of all terms of the form $f^n(t_1, \ldots, t_n)$ for all $n$-place functions $f^n$ occurring in $S$, where $t_j$, $j = 1, \ldots, n$, are members of the set $H_i$. Then each $H_i$ is called the i-level constant set of $S$ and $H_\infty$ is called the Herbrand universe (or H-universe) of $S$, denoted by $H(S)$.

The set of ground atoms of the form $P^n(t_1, \ldots, t_n)$ for all n-ary predicates $P^n$ occuriring in $S$, where $t_1, \ldots, t_n$ are elements of the H-universe of $S$, is called the atom set, or Herbrand base (H-base, for short) of $S$, denoted by $A(S)$.

A ground instance of a clause $C$ of a clause set $S$ is a clause obtained by replacing variables in $C$ by members of H-universe of $S$.

We now consider interpretations over the H-universe. In the following we define a special over the H-universe of $S$, called the H-interpretation of $S$.

**Definition 14.** *Let $S$ be a clause set, $H$ be the H-universe of $S$, and $I=<D,A>$ be an interpretation of $S$. $\mathcal{I}$ is an H-interpretation of $S$ if the following holds:*

- *$D = H$,*
- *Let $c$ be a constant symbol, $c^A = c$,*
- *Let $f$ be a $n$-ary function symbol, $f^A$ maps $(h_1, \ldots, h_n) \in H^n$ to $f(h_1, \ldots, h_n) \in H$*
- *Let $A = \{A_1, \ldots, A_n, \ldots\}$ be the H-base (or atom set) of $S$, H-interpretation $\mathcal{I} = \{m_1, \ldots, m_n, \ldots\}$, where $m_j = A_j$ or $m_j = \neg A_j$.*

Given $I =< D, A >$ interpretation over $D$, an H-interpretation $\mathcal{I} =< H, \mathcal{A} >$ corresponding to $I$ is an H-interpretation that satisfies the following condition:

Let $h_1, \ldots, h_n$ be elements of $H$ and let $m : H \to D$ be a mapping from $H$ to $D$ then $P^{\mathcal{A}}(h_1, \ldots, h_n) = P^A(m(h_1), \ldots, m(h_n))$

Given an Interpretation $I$, we can always find a corresponding $\mathcal{I}$ H-interpretation.

**Theorem 1.** *A clause set $S$ is unsatisfiable iff $S$ is false under all the H-interpretations of $S$.*

Let $S$ be a clause set and $A(S)$ be the H-base of $S$. A *semantic tree* for $S$ is a complete binary tree constructed as follows:

- For each node $N_i$ at the $i^{th}$ level corresponds to an element $A_i$ of $A(S)$, that is, the left edge of $N_i$ is labeled $A_i < \mathsf{W}$, the right edge of $N_i$ is labeled $A_i \geq \mathsf{W}$.
- Conversely, each element of $A(S)$ corresponds to exactly one level in the tree, this means if $A_i \in A(S)$ appears at level $i$ then it must not be at any other levels.

Let $T$ be a semantic tree of a clause set $S$ and $N$ be a node of $T$. We denote $\mathcal{I}(N)$ to be the union of all the sets labeled to the edges of branch of $T$ down to $N$. If there exists an H-interpretation $\mathcal{I}$ in $T$ which contains $\mathcal{I}(N)$, such that $\mathcal{I}(N)$ falsifies some ground instance of $S$, then $S$ is said to be failed at the node $N$. A node $N$ is called a *failure node* of $S$ iff $S$ falsifies at $N$ and $\mathcal{I}(N')$ does not falsify any ground instance of a clause in $S$ for every ancestor node $N'$ of $N$. $N$ is called an *inference node* if all the immediate descendant nodes of N are failure nodes. If every branch in $T$ contains a failure node, cutting off its descendants from $T$, we have $T'$ which is called a *closed tree* of $S$. If the number of nodes in $T'$ is finite, $T'$ is called a finite closed semantic tree.

**Lemma 1.** *Let $S$ be a clause set. Then $S$ is unsatisfiable iff for every semantic tree of $S$, there exists a finite closed tree.*

In the next section we present the inference based on resolution rule for our linguistic logic. Lemma 1 will be used to prove the soundness and completeness of resolution inference rule.

## 4 Resolution

In two-valued logic, when we have a set of formulae $\{A, \neg A\}$ (written as $\{A^{\mathsf{True}}, A^{\mathsf{False}}\}$ in our logic) then the set is said to be contradictory. However in our logic, the degree of contradiction can vary because the truth domain contains more than two elements. Let us consider two sets of formulae $\{A^{\mathsf{VeryTrue}}, A^{\mathsf{VeryFalse}}\}$ and $\{A^{\mathsf{LessTrue}}, A^{\mathsf{LessFalse}}\}$. Then the first set of formulae is "more contradictory" than the second one. Consequently, the notion of reliability is introduced to capture the approximation of linguistic inference.

**Definition 15.** *Let $\alpha$ be an element of $X$ such that $\alpha > \mathsf{W}$ and $C$ be a clause. The clause $C$ with a reliability $\alpha$ is denoted by the pair $(C, \alpha)$.*

The reliability $\alpha$ of a clause set $S = \{C_1, C_2, \ldots, C_n\}$ is defined as follows: $\alpha = \alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n$, where $\alpha_i$ is the reliability of $C_i$ $(i = 1, 2, \ldots, n)$.

A clause $(C_2, \alpha_2)$ is a variant of a clause $(C_1, \alpha_1)$ if $\alpha_1 \neq \alpha_2$ or $C_2$ is equal to $C_1$ except for possibly different variable name.

**Definition 16.** *Given two clauses* $(C_1, \alpha_1)$ *and* $(C_2, \alpha_2)$ *without common variables, where* $C_1 = A^a \vee C_1'$, $C_2 = B^b \vee C_2'$. *Define the linguistic resolution rule as follows:*

$$\frac{(A^a \vee C_1', \alpha_1) \quad (B^b \vee C_2', \alpha_2)}{(C_1'\gamma \vee C_2'\gamma, \alpha_3)}$$

*where* $\gamma$ *is an m.g.u of* $A$ *and* $B$*; and* $a$, $b$, *and* $\alpha_3$ *satisfy the following conditions:* $a \wedge b < \mathsf{W}$, $a \vee b \geq \mathsf{W}$, $\alpha_3 = f(\alpha_1, \alpha_2, a, b)$ *with* $f$ *is a function ensuring that* $\alpha_3 \leq \alpha_1$, *and* $\alpha_3 \leq \alpha_2$.
$(C_1'\gamma \vee C_2'\gamma, \alpha_3)$ *is a binary resolvent of* $(C_1, \alpha_1)$ *and* $(C_2, \alpha_2)$. *The literals* $A^a$ *and* $B^b$ *are called literals resolved upon.*

In Def. 16, $\alpha_3$ is defined so as to be smaller or equal to both $\alpha_1$ and $\alpha_2$. In fact, the obtained clause is less reliable than original clauses. The function $f$ is defined as following:

$$\alpha_3 = f(\alpha_1, \alpha_2, a, b) = \alpha_1 \wedge \alpha_2 \wedge (\neg(a \wedge b)) \wedge (a \vee b) \tag{1}$$

Obviously, $\alpha_1, \alpha_2 \geq \mathsf{W}$, and $\alpha_3$ depends on $a, b$. Additionally, $a \wedge b < \mathsf{W}$ implies $\neg(a \wedge b) > \mathsf{W}$. Moreover, $(a \vee b) \geq \mathsf{W}$. Then, by Formula (1), we have $\alpha_3 \geq \mathsf{W}$.

An inference is sound if its conclusion is a logical consequence of its premises. That is, for any interpretation I, if the truth values of all premises are greater than $\mathsf{W}$, the truth value of the conclusion must be greater than $\mathsf{W}$.

**Definition 17.** *A resolvent of clauses* $C_1$ *and* $C_2$ *is a binary resolvent of factors of* $C_1$ *and* $C_2$, *respectively.*

**Definition 18.** *Let* $S$ *be a clause set. A resolution derivation is a sequence of the form* $S_0, \ldots, S_i, \ldots$, *where*

- $S_0 = S$, *and*
- $S_{i+1} = S_i \cup \{(C, \alpha)\}$, *where* $(C, \alpha)$ *is the conclusion of a resolution inference with premises* $S_i$ *based on resolution rule in Def. 16 and* $(C, \alpha) \notin S_i$.

We find that resolution derivation $S_0, \ldots, S_i, \ldots$ is infinite because the set of assignments and the set of semantic values are infinite. However, if the original clause set $S$ is unsatisfiable, the sequence $S_i$ always derives an empty clause $\square$. The soundness and completeness of resolution derivation is shown by the following theorem:

**Theorem 2.** *Let* $S$ *be a clause set,* $S_0, \ldots, S_i, \ldots$ *be a resolution derivation.* $S$ *is unsatisfiable iff there exists* $S_i$ *containing the empty clause* $\square$.

A *resolution proof* of a clause $C$ from a set of clauses $S$ consists of repeated application of the resolution rule to derive the clause $C$ from the set $S$. If $C$ is the empty clause then the proof is called a *resolution refutation*. We represent resolution proofs as *resolution trees* where a node is labeled with a clause. The root node without child node is labeled with the conclusion clause. All nodes

with no parent node are labeled with clauses from the initial clause set $S$. All other nodes must have two parents and are labeled with a clause $C$ such that

$$\frac{C_1 \quad C_2}{C}$$

where $C_1, C_2$ are the labels of the two parent nodes. If $\mathsf{RT}$ is a resolution tree representing the proof of a clause with reliability $(C, \alpha)$, then we say that $\mathsf{RT}$ has the reliability $\alpha$.

*Example 1.* Let $AX = (X, \{\mathsf{False}, \mathsf{True}, \}, H, \leq$ be a *lin-HA* where $H^+ = \{\mathsf{V}, \mathsf{M}\}$ and $H^- = \{\mathsf{P}, \mathsf{L}\}$ *(V=Very, M=More, P=Possible, L=Less)*; Consider the clause set after transforming into CNF as following:

1. $A(x)^{\mathsf{MFalse}} \vee B(z)^{\mathsf{MFalse}} \vee C(x)^{\mathsf{PTrue}}$
2. $C(y)^{\mathsf{MFalse}} \vee D(y)^{\mathsf{VMTrue}}$
3. $C(t)^{\mathsf{VVTrue}} \vee E(t, f(t))^{\mathsf{MFalse}}$
4. $E(a, u)^{\mathsf{True}}$
5. $A(a)^{\mathsf{VTrue}}$
6. $B(a)^{\mathsf{LTrue}}$
7. $D(a)^{\mathsf{MFalse}}$

where $a, b$ are constant symbols; $t, x, y, u, z$ are variables. At the beginning, each clause is assigned to the highest reliability $\top$. We have a resolution proof as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{(A(x)^{\mathsf{MFalse}} \vee B(z)^{\mathsf{MFalse}} \vee C(x)^{\mathsf{PTrue}}, \top) \quad (A(a)^{\mathsf{VTrue}}, \top)}{(B(z)^{\mathsf{MFalse}} \vee C(a)^{\mathsf{PTrue}}, \mathsf{MTrue})}[a/x] \quad (B(a)^{\mathsf{LTrue}}, \top)}{(C(a)^{\mathsf{PTrue}}, \mathsf{LTrue})}[a/z] \quad (C(y)^{\mathsf{MFalse}} \vee D(y)^{\mathsf{VMTrue}}, \top)}{(D(a)^{\mathsf{VMTrue}}, \mathsf{LTrue})}[a/y] \quad (D(a)^{\mathsf{MFalse}}, \top)}{(\square, \mathsf{LTrue})}$$

## 5 Conclusion

We have presented syntax and semantics of our linguistic first order logic system. We based on linear symmetrical hedge algebra to model the truth value domain. To capture the approximate of inference in nature language, each clause in our logic is associated with a reliability. We introduced an inference rule with a reliability which ensures that the reliability of the inferred clause is less than or equal to those of the premise clauses. Based on the algebraic structure of linear symmetrical hedge algebra, resolution in linguistic first order logic will contribute to automated reasoning on linguistic information. It would be worth investigating how to extend our result to other hedge algebra structures and to other automated reasoning methods.

**Acknowledgement.** The authors would like to thank the National Foundation of Science and Technology Development for its support under Grant 102.04-2013.21.

## References

[1] Chang, C.-L., Lee, R.C.-T.: Symbolic Logic and Mechanical Theorem Proving, 1st edn. Academic Press, Inc., Orlando (1997)
[2] Ebrahim, R.: Fuzzy logic programming. Fuzzy Sets and Systems 117(2), 215–230 (2001)
[3] Esteva, F., Godo, L., Noguera, C.: A logical approach to fuzzy truth hedges. Information Sciences 232, 366–385 (2013)
[4] Ho, N.C., Wechler, W.: Extended hedge algebras and their application to fuzzy logic. Fuzzy Sets and Systems 52(3), 259 (1992)
[5] Háek, P.: On very true. Fuzzy Sets and Systems 124(3), 329 (2001)
[6] Klement, E.P.: Some mathematical aspects of fuzzy sets: triangular norms, fuzzy logics, and generalized measures. Fuzzy Sets Syst. 90(2), 133–140 (1997)
[7] Le, V.H., Liu, F., Tran, D.K.: Fuzzy linguistic logic programming and its applications. Theory Pract. Log. Program. 9(3), 309–341 (2009)
[8] Lee, R.C.T.: Fuzzy logic and the resolution principle. J. ACM 19(1), 109–119 (1972)
[9] Mondal, B., Raha, S.: Approximate reasoning in fuzzy resolution. In: 2012 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS), pp. 1–6 (August 2012)
[10] Nguyen, C.H., Wechler, W.: Hedge Algebras: An Algebraic Approach in Struture of Sets of Linguistic Truth Values. Fuzzy Sets and Syst. 35, 281–293 (1990)
[11] Nguyen, T.-M.-T., Vu, V.-T., Doan, T.-V., Tran, D.-K.: Resolution in linguistic propositional logic based on linear symmetrical hedge algebra. In: Huynh, V.N., Denoeux, T., Tran, D.H., Le, A.C., Pham, B.S. (eds.) KSE 2013, Part I. AISC, vol. 244, pp. 337–350. Springer, Heidelberg (2014)
[12] Phuong, L.A., Khang, T.D.: A deductive method in linguistic reasoning. In: 2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering (URKE), pp. 137–140 (2012)
[13] Phuong, L.A., Khang, T.D.: Linguistic reasoning based on generalized modus ponens with linguistic modifiers and hedge moving rules. In: 2012 International Conference on Fuzzy Theory and it's Applications (iFUZZY), pp. 82–86 (2012)
[14] Robinson, J.A.: A machine-oriented logic based on the resolution principle. J. ACM 12(1), 23–41 (1965)
[15] Shen, Z., Ding, L., Mukaidono, M.: Fuzzy resolution principle. In: Proceedings of the Eighteenth International Symposium on Multiple-Valued Logic, pp. 210–215 (1988)
[16] Smutná-Hliněná, D., Vojtáš, P.: Graded many-valued resolution with aggregation. Fuzzy Sets and Systems 143(1), 157–168 (2004)
[17] Tran, D.-K., Vu, V.-T., Doan, T.-V., Nguyen, M.-T.: Fuzzy linguistic propositional logic based on refined hedge algebra. In: 2013 IEEE International Conference on Fuzzy Systems (FUZZ), pp. 1–8 (2013)
[18] Vojtás, P.: Fuzzy logic programming. Fuzzy Sets and Systems 124(3), 361–370 (2001)
[19] Vychodil, V.: Truth-depressing hedges and bl-logic. Fuzzy Sets and Systems 157(15), 2074 (2006)
[20] Weigert, T.J., Tsai, J.-P., Liu, X.: Fuzzy operator logic and fuzzy resolution. J. Autom. Reasoning 10(1), 59–78 (1993)
[21] Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)

# A Fuzzy Set Based Evaluation of Suppliers on Delivery, Front Office Quality and Value-Added Services

Margaret F. Shipley, Gary L. Stading, and Jonathan Davis

University of Houston Downtown
326 North Main Street
Houston, Texas 77002
shipleym@uhd.edu, stadingg@uhd.edu, davisg@uhd.edu

**Abstract.** Fuzzy probabilities are used in an algorithmic process to address the ambiguity and uncertainty in supplier selection. Supplier selection is receiving increased focus in supply chain management (SCM) and was the impetus of a survey sent to 3000 companies that deal with an industry-dominated seven suppliers. This study focuses on three criteria, each having four attributes; delivery, front-office quality, and value-added services. The respondent data are partitioned, the algorithm is applied to the twelve aspects of the criteria using a spreadsheet program, the results are analyzed, and discussion is provided of a weighted scoring approach to rank order the suppliers.

**Keywords:** Fuzzy Set Theory, Fuzzy Probability, Supplier Selection.

## 1 Introduction

Supplier selection decisions are receiving increased focus in industry, and the associated academic papers have no shortage of ideas about how selection decisions can be made. Alternatives range from simple expert opinion methods that are not reliably repeatable to quantitative methods that may not account for various criteria associated with standards for supplier performance. Even where quantitative supplier performance data are readily available, subjective judgment of qualitative performance metrics must be provided by a variety of sources, including senior management ([1],[2]), experts in the field ([3],[4],[5]) and even a project's team members [6]. However, subjective judgments offered in terms of linguistic variables provide a degree of uncertainty and input ambiguity into the decision. Customer demands are generally uncertain and supplier evaluation, selection and coordination lead to various strategies to manage supplier relationships ([7],[8]). Fuzzy logic has been recognized as an important tool in the analysis of uncertainty in decision making situations, including supply chain management (SCM).

Lui [9], proposed a fuzzy model for partial backordering models in 1999. Five years later inventory discounting considered the buyer-seller relationships [10], and location aspects for inventory control became fuzzy considerations [11]. Supply chain decisions for integrated just-in-time inventory systems recognized the fuzzy nature of annual demand and production rates as being no longer statistically based. The assumption

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 355–365, 2014.

of known annual demand was considered by the authors to be unrealistic such that the proposed model included fuzzy annual demand and/or production rate, employing the signed distance, a ranking method for fuzzy numbers, to estimate fuzzy total cost of the JIT production in today's supply chain environment. A fuzzy-set based method derived the optimal buyer's quantity and number of lots from the vendor [12].

Fuzzy programming contributed to the following: optimal product mix based on ABC analysis [13]: fuzzy multi-objective linear programming minimized total production and transportation costs; the number of rejected items and total delivery time as related to labor and budget constraints [14]; and fuzzy goal programming considered supply chain management from the perspective of activity-based costing with mathematically derived optimization for evaluating performance of the value-chain relationship [15]. Manufacturing processes as related to business logistics looked at the data itself as fuzzy in Quality Function Deployment's relationship to customer service [16]. The pursuit of goals such as quality further led to attempts to balance production process capacities of assembly lines. Fuzzy goals were used as a tool for measuring, displaying and controlling industrial process variables [13].

Considering different quality standards in a supply chain network, Chan et al. used a fuzzy neural approach to suggest adjustments of product quantity from various suppliers [17]. The Fuzzy Suitability Index (FSI) aggregated rankings and multiplied, by weight, each criterion [18]. With the same goal of ranking suppliers according to performance, a method was proposed whereby n decision makers evaluated the performance of m suppliers in k criteria, rating the importance of the k criteria in linguistic terms. Aggregation of the fuzzy expressions for importance weights, and a fuzzy preference index led to rank ordering of the suppliers [19].

Supplier selection was developed from a rule-based perspective. The approach selected was fuzzy associated rule mining from the database for supplier assessment [20]. Sevkli [21] in his comparison of a recognized crisp ELECTRE model versus a fuzzy ELECTRE model, reached the conclusion that using fuzzy sets for multi-criteria supplier selection decisions is superior.

## 2 Fuzzy Supplier Selection Model

Fuzzy logic addresses the ambiguity of data and uncertainty in this decision making situation, where a fuzzy subset $A$ of a set $X$ is a function of $X$ into [0,1]. For a brief foundation in the basics, see ([22], [23],[24]). While a new class of implication operators has been proposed [25], the more traditionally utilized fuzzy operations are used in this research. $A$ and $B$ denote two fuzzy sets, so the intersection, union, and complement are defined by:

$A \cap B = \Sigma \gamma_i / x_i$, where $\gamma_i = \text{Min} \{\alpha_i, \beta_i\}$;
$A \cup B = \Sigma \gamma_i / x_i$, where $\gamma_i = \text{Max} \{\alpha_i, \beta_i\}$;
$\neg A = \Sigma \gamma_i / x_i$, where $\gamma_i = 1 - \alpha_i$;
and it is assumed that $B = \Sigma \beta_i / x_i$ ([26],[27],[28],[29])

Extension principles ([30],[23],[31]) often guide the computations when dealing with fuzzy sets. Letting $f$ be a function from $X$ into $Y$, with $Y$ as any set and $A$ as above, then $f$ can be extended to fuzzy subsets of $X$ by:

$$f(A) = \Sigma_y \, u_{f(A)}(y) \,/\, y, \text{ where } u_{f(A)}(y) = \mathrm{Max}_{x \varepsilon f^{-1}(y)} \, A(x)\mathrm{T}$$

thus, $f(A)$ is a fuzzy subset of $Y$. In particular, if $f$ is a mapping from a Cartesian product such as $X \times Y$ to any set, $Z$, then $f$ can be extended to objects of the form $(A,B)$ where $A$ and $B$ are fuzzy subsets of $X$ and $Y$ by:

$f(A,B) = \Sigma \, u_{f(A,B)}(z) \,/\, z$, where $u_{f(A,B)}(z) = \mathrm{Max}_{(x,y) \varepsilon f^{-1}(z)} \, \mathrm{Min}\{A(x), B(x)\}$.

A fuzzy set P whose elements all lie on the interval [0,1] can be expressed as a fuzzy probability. Consider a set of $n$ fuzzy probabilities each having $r$ elements, $a_i = \sum_{j=1}^{r} \alpha_{ij} \,/\, a_{ij}$ for $i = 1, 2, \ldots, n$, where $\alpha_{ij}$ denotes the degree of belief that a possible value of $a_i$ is $a_{ij}$. Then $(a_1, a_2, \ldots a_n)$ constitutes a finite fuzzy probability distribution if and only if there are $n$-tuples $a_i$, $i = 1,2,\ldots,n$ such that $\sum_{i=1}^{n} a_i = 1$.

To qualify as a finite fuzzy probability distribution, each fuzzy probability in the distribution must have the same number of elements (some of the $a$'s may be zero), and these elements should be ordered in the sense that the sum of the elements in each specific position must equal one. So the $n$-tuples $(a_{ij})$, $i$=1,2,...,$n$ form probability distributions in the crisp sense. This type of probability distribution can be transformed such that the resulting distribution has entropy at least as great as the original [32].

A version of fuzzy expected values was first used when Zebda [31] defined $Q_{ijk} = \Sigma \, \alpha_{ijk} \,/\, a_k$ as the fuzzy probability that from State $i$ and making Decision $j$, reach State $k$. Associated with this are fuzzy benefits $B_{ijk}$ where $B_{ijk} = \Sigma \, ß_{ijk} \,/\, b_k'$

Then the averaged benefit is defined by $E(B_{ijk}) = \Sigma c_{ij\ell} \,/\, b_\ell$ where:

$c_{ij\ell} = \boldsymbol{Max}_{(a_1,\ldots a_p, b_1', \ldots b_p') \varepsilon f^{-1} b_\ell k} \, \boldsymbol{Min}\left(\alpha_{ijk}, ß_{ijk}\right)$

for $b_\ell = \Sigma_k \, a_x b_x$ if $\Sigma_k \, a_x = 1$ and 0 otherwise.

Here, $f(a_1,\ldots,a_p,b_1',\ldots,b_p') = \Sigma \, a_x b_x'$

## 2.1 Algorithm

The algorithm preserves information during the process of computing and evaluating fuzzy probabilities until a final weighted model collapses the results into an objective score.

0. Randomly partition the criteria data set into $\ell$ subsets of equal size.
1. For each attribute $\phi$ of each supplier $v$, subjectively assign scores $s_{\phi kv}$. The supplier rating $(s_{\phi kv})$ is then given by the equation $s_{\phi kv} = \sum \, \tau_{\phi k} \,/\, s_{\phi k}$ for all $v$ where $\tau_{\phi k} = 1$ ($v$=1,2,…,$m$; $k$=1,2,…,$n$; and $1 < \phi < x$).
2. Define the fuzzy expected value, $Q_{\phi kv}$, for each attribute $\phi$ of each $v$ in terms of each $s_{\phi kv}$ as $Q_{\phi kjv} = \sum \, \alpha_{\phi kjv} \,/\, a_{\phi kjv}$ for all $s_{\phi kjv}$, where each $\alpha_{\phi kjv}$ represents belief in the probability $a_{\phi kjv}$ that $v$ will be scored $s_{\phi kjv}$ ($v$ =1,2,…,$m$; $k$=1,2,…,$n$; $1 < \phi < x$ and $j$=1,2,…,$\ell$).
3. Group the probabilities $a_{\phi kjv}$ into combinations $\varphi_{\phi v}$ such that $\sum a_{\phi kjv} = 1$ for some set $H$ of $k$'s. $a_{\phi kjv} = 0$ for $k \notin H$.

4. Across all partitions $\ell$, compute $b_{\phi v} = \{\sum a_{\phi kjv} s_{\phi kjv}$ if $\sum a_{\phi kjv} = 1$, otherwise 0 ($k = 1,2,\ldots,r$; $j$=1,2,…,$\ell$ and $p$ = the distinct number of $\sum a_{\phi kjv} = 1$; $1< \ell \leq p$).
5. For all $\alpha_{\phi kjv} \neq 0$ find $c_{\phi v}$ = Min $\{ \tau_{\phi kjv}, \alpha_{\phi kjv}\}$, where $c_{\phi v}$ is the degree of belief that the expected value is $b_{\phi v}$.
6. Defuzzify the expected value for each attribute $\phi$ to find E($s_{\phi v}$) = $\sum c_{\phi v} b_{\phi v} / \sum c_{\phi v}$.

# 3 Application

The application presents a real-world supplier selection decision-making problem based upon: 1) generation of data from a survey of purchasing professionals, and 2) partitioning of the resulting data to fit the algorithm detailed above.

## 3.1 Example Data

A survey was distributed to about 3,000 companies in Texas that purchase semiconductors, passives, RF/microwaves, connectors and interconnects, and electromechanical devices from a small set of dominant suppliers. Representative industries included automotive, communications, contract design/engineering, power/electrical, medical/dental, computer, manufacturing, and military/aerospace. The survey queried each customer's number of years of activity in the industry in designated ranges from less than two to 21 or more. Customers dealt with multiple suppliers and specified their firm's annual sales revenue as under $5,000,000 to over $2,000,000,000. With 412 surveys received, the response rate was slightly under 15%.

For model application purposes, the survey provided performance measurements on each supplier, as well as measures of the importance of each criterion to the customer and the customer's level of belief explicitly tied to the company's annual amount of business conducted with the targeted group of suppliers. Survey questions relating directly to the importance of this fuzzy supplier selection application included a query of the amount of money the customer spends on electronic components in a year. These ranges were: <$100,000; $100,000..$499,999; $500,000..$999,999; $1,000,000..$9,999,999; $10,000,000..$24,999,999; and >$25,000,000. These ranges were used to identify a firm's level of activity with the suppliers in question and, therefore, its expected level of confidence (interpreted as r belief) in its assessments (see Table 2).

Survey questions emphasized different aspects of quality by which a customer would select a supplier. Delivery specific questions related to on-time performance, availability of inventory, shipping accuracy and return authorization process. Front office quality was assessed based on quote completeness, credit and payment terms, cancelled/non-returnable letters, and contract processing. Finally, value-added services dealt with customized processing, knowledgeable specialists, technical design services, e-business services and sales management. Price was assumed to be competitively based for the seven suppliers/ vendors evaluated by the industry respondents. However, one supplier, Kent, was removed from the set due to: a) low survey responses compared to the other suppliers, and b) no longer existing as an independent company, having been acquired by Avnet after the survey was conducted.

An extensive study by Simpson et al. [33] determined the highest number of forms and percentage of all forms containing Price, Delivery, Quality (PDQ) with other selection criteria. To validate our survey's criteria relevance, a comparison was made as follows:

**Table 1.** Comparison of Survey Attributes to Comprehensive Survey of Relevance to Supplier Evaluation

| Survey Category | Industry Question | Simpson et al. | % of Forms using |
|---|---|---|---|
| | | | |
| Delivery | On-time delivery | Delivery timeliness | 61.9 |
| | Availability of inventory | Inventory accuracy<br>Fill Rate | 15.5<br>15.5 |
| | Shipping accuracy | Accurate delivery<br>Inspection Prior to Shipping | 32.1<br>27.4 |
| | Return material authorization | Return procedures<br>Complaint Handling Process | 20.2<br>33.3 |
| | | | |
| **Front Office Quality** | Quote completeness & turnaround | Quality documentation<br>Prompt ordering process<br>Timely ordering | 48.8<br>28.6<br>16.7 |
| | Credit and payment terms | Payment process | 10.7 |
| | Non-cancellable-return letters | Corrective/preventative measures | 54.8 |
| | Contract processing | Customer/PO requirements met<br>Accurate invoicing | 78.6<br>20.2 |
| | | | |
| **Value Added Services** | Knowledge specialists adding value | Staff problem solver<br>Staff expertise | 11.9<br>20.2 |
| | Technical design services | Technical assistance | 32.1 |
| | e-business service | EDI capability<br>Inventory mgt. system<br>Inventory tracking | 14.3<br>35.7<br>35.7 |
| | Sales & sales management support | Quality management<br>Staff responsive | 54.8<br>21.4 |

Given that real-world collection of survey data is, to varying degrees, imperfect, a threshold was established below which respondent data would be considered incomplete and removed. The resulting dataset left a pool of 150 useful responses to be applied to the fuzzy algorithm. The remaining survey responses were randomly partitioned into two sets of 75 responses each in accordance with Step 0 of the model algorithm.

## 4 Modeling Process

By Step 1 of the algorithm, $\phi$ =1,2,3,4 attributes as defined above. Each of the four attributes is subjectively assigned a score by the respondent for each of the six suppliers (m=6), equating to Poor, Below Average, Average, Above Average and Excellent (n=5). Supplier rating $s\phi\nu$ is then given by the equation $s\phi\nu = \sum\tau\phi k\nu / s\phi k\nu$ for each supplier, $\nu$, and, by Step 2, the fuzzy probability $Q\phi kj\nu$, for each attribute of $\nu$ in terms of $s\phi kj\nu$ is $Q\phi kj\nu = \sum \alpha\phi kj\nu / a\phi kj\nu$ for all $s\phi\nu$. Each $\alpha\phi kj\nu$ represents belief in the probability $a\phi kj\nu$ that $\nu$ will perform to the level of the assigned score $s\phi\nu$ (k=1,2,…,5; $\nu$ =1,2,…,6; $\phi$ =1,2,3,4; and j=1,2).

Table 2 shows the scoring of respondent belief as proportional to total possible spending.

**Table 2.** Respondent Belief Associated with Spending

| Spending | Degree of belief |
|---|---|
| < $100,000 | 0.0020 |
| < $500,000 | 0.0100 |
| < $1,000,000 | 0.0300 |
| < $10,000,000 | 0.2000 |
| < $25,000,000 | 0.7000 |
| > $25,000,000 | 1.0000 |

## 5 Results

The fuzzy probabilities from the respondents for the six suppliers are found by Step 3 of the algorithm and result in four fuzzy probability (summation to 1.0) combinations for $\nu_3$; 8 combinations for $\nu_4$; 5 combinations for $\nu_5$ and 2 combinations for $\nu_6$. Using spreadsheet computerization of the algorithmic process for Steps 4, 5, and 6 using six suppliers:

**Table 3.** Results for Delivery Category Attributes

| Supplier | On-Time Delivery | Availability of Inventory | Shipping Accuracy | Return Authorization | Overall Average |
|---|---|---|---|---|---|
| $v_1$: **Arrow** | 3.80 | 3.77 | 3.93 | 3.80 | 3.82 |
| $v_2$: **Avnet** | 3.69 | 3.63 | 3.90 | 3.68 | 3.72 |
| $v_3$:**Future** | 3.71 | 3.60 | 3.81 | 3.30 | 3.61 |
| $v_4$:**Insight** | 3.65 | 3.33 | 3.83 | 3.66 | 3.62 |
| $v_5$:**Pioneer** | 3.57 | 3.27 | 3.70 | 3.61 | 3.54 |
| $v_6$: **TTI** | **3.89** | 3.71 | **4.06** | 3.68 | **3.83** |

**Table 4.** Results for Front Office Quality Attributes

| Supplier | Quote Com-pleteness | Credit/PMT Terms | Non-Return Letters | Contract Terms | Overall Average |
|---|---|---|---|---|---|
| $v_1$: **Arrow** | **3.81** | **4.02** | 3.60 | **3.71** | **3.79** |
| $v_2$: **Avnet** | 3.49 | 3.92 | 3.32 | 3.60 | 3.58 |
| $v_3$:**Future** | 3.53 | 3.81 | 3.34 | 3.42 | 3.53 |
| $v_4$: **Insight** | 3.55 | 3.68 | 3.36 | 3.51 | 3.53 |
| $v_5$: **Pioneer** | 3.68 | 3.74 | 3.51 | 3.41 | 3.58 |
| $v_6$: **TTI** | 3.79 | 3.81 | **3.62** | 3.54 | 3.69 |

**Table 5.** Results for Value-Added Services Attributes

| Supplier | Knowledge Specialists | Technical Design Help | e-Business Services | Sales/Mgt. Support | Overall Average |
|---|---|---|---|---|---|
| $v_1$: **Arrow** | **3.90** | **3.51** | 3.56 | **3.80** | **3.69** |
| $v_2$: **Avnet** | 3.64 | 3.46 | 3.41 | 3.57 | 3.52 |
| $v_3$: **Future** | 3.50 | 3.39 | 3.02 | 3.46 | 3.34 |
| $v_4$:**Insight** | 3.46 | 3.37 | 2.98 | 3.45 | 3.31 |
| $v_5$:**Pioneer** | 3.57 | 3.25 | 3.17 | 3.47 | 3.37 |
| $v_6$: **TTI** | 3.63 | 3.28 | **4.19** | 3.66 | 3.44 |

## 6 Discussion

Using an individual criteria approach based on simple averaging, TTI and Arrow perform best on Delivery attributes (3.83, 3.82, respectively); Arrow is the best for Front Office Quality (3.79), and Arrow is best for Value-Added Services (3.69). Looking more closely into the attributes comprising the criterion category, Arrow outperforms all others in Availability of Inventory (3.77) and Return Authorization (3.80) while TTI outperforms all others in On-time delivery (3.89) and Shipping Accuracy (4.06). For Front Office attributes, TTI exceeds Arrow only in Non-Return letters (3.62). Similarly, TTI outperforms Arrow only on the e-Business Services (4.19) attribute of Value-Added Services. Of note, is although Arrow scores higher on several attributes, its score on all twelve of these is only higher than a 4 on Credit/PMT terms (4.02) while TTI has the overall highest score (4.19) on e-Business Services and another score that exceeds 4 on Shipping Accuracy (4.06).

Considering a simple rank ordering process for each criterion category and letting D denote Delivery, FO denote Front Office Quality, and VA denote Value Added, we can see where each supplier ranked in comparison with all others:

**Table 6.** Ranking of Suppliers on each Decision Criterion Category

| *Supplier* | *First* | *Second* | *Third* | *Fourth* | *Fifth* | *Sixth* |
|---|---|---|---|---|---|---|
| **Arrow** | D(1), FO(3), VA(3) | D(3) FO(1), VA(1) | | | | |
| **Avnet** | | FO(2), VA(2) | D(3), FO(1), VA(2) | D(1) | | FO(2) |
| **Future** | | | D(1), VA(1) | D(1) | D(1), FO(3), VA(3) | D(1) |
| **Insight** | | | | D(2), FO(3), VA(1) | D(2) | FO(1), VA(3) |
| **Pioneer** | | | FO(2) | VA(3) | D(1), FO(1) | D(3), FO(1), VA(1) |
| **TTI** | D(3), FO(1), VA(1) | D(1), FO(1), VA(1) | FO(2), VA(1) | | VA(1) | |

Overall rankings based on the number of times ranked first, second, to sixth show Arrow as the first place winner over TTI, but Avnet almost ties with Arrow for second place by outperforming in both Front-office Quality and Value Added Services. Avnet also garners the most third place rankings. The other suppliers generally compete for fourth to sixth place with Insight being better in all areas except Value Added where Pioneer performs better. Assigning points for first as 11, second as 9, etc. then the overall ranking would be: Arrow, TTI and Avnet, with the others basically tied.

If, however, the buyer is more concerned with a particular criterion, then weights of relative importance could be applied. Simpson's work [33] provides a list of criteria and the percentage of time these are observed as being on the supplier evaluation forms. We have by Table 1 correlated our attributes to those of the Simpson [33] study and by extension to the criteria/categories of Delivery, Front-Office Quality and Value-Added Services. Assuming that the attributes expressed in the Simpson study [33] are as we have equated them to our criteria, the first two attributes of importance are within Front-office Quality; contract processing (98.8%) followed by quote completeness and turnaround (94.1%). The leading supplier in both regards is Arrow. Third and fourth on the list of most frequently used on evaluation of supplier forms are in Value-Added Services; specifically e-business service at 85.7% and sales management support at 76.2%. TTI ranks higher in the former and Arrow ranks higher in the latter. The next three most likely aspects for supplier evaluation per Simpson [33] and our correlated categories show TTI to be the best. Assuming a threshold of 55% of the companies using the decision criteria, TTI would be selected based upon on-time delivery, shipping accuracy, and non-cancellable-return letters. If the threshold is set lower, Arrow captures all other attributes considered in this study.

## 7 Conclusions

As the Discussion indicates, this study has no solution. Instead, the focus was on development and analysis of the process. Following the fuzzy probability based algorithmic process is easy for simple example illustrations. This study expands beyond explanation and illustration purposes in an attempt to put a large set of data into the model. The results are basically brute-forced through Excel looking at all combinatorial fuzzy probably summations of 1.0 or greater. So, use of the algorithm and an Excel spreadsheet approach has been shown to be able to resolve this supplier selection decision making problem.

There is, however, a need to determine these probability combinations as input into the spreadsheet. The authors are, thus, developing an interface to the spreadsheet that will determine all combinations across all the qualitatively defined score values and to use only those that satisfy Step 3 of the algorithm.

The ability to translate a simple survey into fuzzy sets is a plus for this study. Typically, gathered data can be translated by the process shown. However, designing a survey that more easily determines belief and membership functions would be preferable. Still, the Excel process is simple to use and easy transference of data is useful for business decisions.

## References

1. Ghodsypour, S.H., O'Brien, C.: A decision support system for supplier selection using an integrated analytic hierarchy process and linear programming. International Journal of Production Economics, 56-57, 99-122 (1998)
2. Verma, R., Pullman, M.E.: An analysis of the supplier selection process. Omega 26, 739–750 (1998)

3. Mandal, A., Deshmukh, S.G.: Vendor selection using interpretive structural modeling (ISM). International Journal of Operations and Production Management 14, 52–59 (1994)
4. Cheng, E.W.L., Li, H.: Information priority-setting for better resource allocation using analytic hierarchy process (AHP). Information Management and Computer Security 9, 61–70 (2001)
5. Humphreys, P.K., Shiu, W.K., Chan, F.T.S.: Collaborative buyer-supplier relationships in Hong Kong manufacturing firms. Supply Chain Management: an International Journal 6, 152–162 (2001)
6. Ragatz, G.L., Handfield, R.B., Scannell, T.V.: Success factors for integrating suppliers into new product development. Journal of Product Innovation Management 14, 190–202 (1997)
7. Chan, F.T.S.: Interactive selection model for supplier selection process: an analytical hierarchy process approach. International Journal of Production Research 41(15), 3549–3579 (2003)
8. Deng, S.-J., Elmaghraby, W.: Supplier selection via tournaments. Production and Operations Management 14(2), 252–267 (2005)
9. Liu, B.: Fuzzy criterion models for inventory systems with partial backorders. Annals of Operations Research 87, 117–126 (1999)
10. Das, K., Roy, T.K., Maiti, M.: Buyer–seller fuzzy inventory model for a deteriorating item with discount. International Journal of Systems Science 35, 457–466 (2004)
11. Usenik, J., Bogataj, M.: A Fuzzy Set Approach for a Location-Inventory Model. Transportation Planning and Technology 28, 447–464 (2005)
12. Pan, J.C.H., Yang, M.F.: Integrated inventory models with fuzzy annual demand and fuzzy production rate in a supply chain. International Journal of Production Research 46, 753–770 (2007)
13. Kara, Y., Gökçen, H., Atasagun, Y.: Balancing parallel assembly lines with precise and fuzzy goals. International Journal of Production Research 48, 1685–1703 (2009)
14. Liang, T.F.: Integrating production-transportation planning decision with fuzzy multiple goals in supply chains. International Journal of Production Research 46, 1477–1494 (2007)
15. Tsai, W.H., Hung, S.-J.: A fuzzy goal programming approach for green supply chain optimisation under activity-based costing and performance evaluation with a value-chain structure. International Journal of Production Research 47, 4991–5017 (2009)
16. Shu, M.-H., Wu, H.-C.: Measuring the manufacturing process yield based on fuzzy data. International Journal of Production Research 48, 1627–1638 (2009)
17. Chan, F.T.S., Kumar, N., Tiwari, M.K., Lau, H.C.W., Choy, K.L.: Global supplier selection: a fuzzy-AHP approach. International Journal of Production Research 46, 3825–3857 (2008)
18. Bevilacqua, M., Petroni, A.: From traditional purchasing to supplier management: a fuzzy logic-based approach to supplier selection. International Journal of Logistics 5, 235–255 (2002)
19. Bayrak, M.Y., Çelebi, N., Taşkin, H.: A fuzzy approach method for supplier selection. Production Planning & Control 18, 54–63 (2007)
20. Jain, V., Wadhwa, S., Deshmukh, S.G.: Supplier selection using fuzzy association rules mining approach. International Journal of Production Research 45, 1323–1353 (2006)
21. Sevkli, M.: An application of the fuzzy ELECTRE method for supplier selection. International Journal of Production Research 48, 3393–3405 (2009)
22. Zadeh, L.A.: Generalized theory of uncertainty (GTU)—principal concepts and ideas. Computational Statistics & Data Analysis 51, 15–46 (2006)
23. Dubois, D., Prade, H.: Systems of linear fuzzy constraints. Fuzzy Sets and Systems 3, 37–48 (1980)

24. Freeling, A.N.: Fuzzy sets and decision analysis. IEEE Transactions on Systems, Man and Cybernetics 10, 341–354 (1980)
25. Yager, R.R.: On some new classes of implication operators and their role in approximate reasoning. Information Sciences 167, 193–216 (2004)
26. Kaufmann, A., Gupta, M.: An introduction to fuzzy sets arithmetic. Nosfrand Reinhold Co., New York (1985)
27. Klir, G.J., Folger, T.A.: Fuzzy sets, uncertainty, and information (1988)
28. Zadeh, L.: Fuzzy Sets. Information and Control 8, 338–353 (1965)
29. Zadeh, L.A.: Fuzzy logic and approximate reasoning. Synthese 30, 407–428 (1975)
30. Gupta, M.M., Ragade, R.K., Yager, R.R.: Advances in fuzzy set theory and applications. North-Holland Pub. Co. (1979)
31. Zebda, A.: The investigation of cost variances: A fuzzy set theory approach. Decision Sciences 15, 359–388 (1984)
32. Yager, R.R., Kreinovich, V.: Entropy conserving probability transforms and the entailment principle. Fuzzy Sets and Systems 158, 1397–1405 (2007)
33. Simpson, P.M., Siguaw, J.A., White, S.C.: Measuring the Performance of Suppliers: An Analysis of Evaluation Processes. Journal of Supply Chain Management 38, 29–41 (2002)

# A Fuzzy Rulebase Approach to Remap Gridded Spatial Data: Initial Observations

Jörg Verstraete

Systems Research Institute - Department of Computer Modelling
Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
jorg.verstraete@ibspan.waw.pl
http://ibspan.waw.pl

**Abstract.** In many fields of research where different gridded spatial data needs to be processed, the grids do not align properly. This can be for a multitude of reasons, and it complicates drawing conclusions and further processing the data; it requires one grid to be transformed to match the other grid. In this article, we present the first results of a completely new approach to transforming data that are represented in one grid, to have it match a given target grid. The approach uses techniques from artificial intelligence and simulates an intelligent reasoning on how the grid can be transformed, using additionally available information to estimate the underlying distribution. The article describes the algorithm, and results on artificial datasets are discussed.

## 1 Introduction

### 1.1 Problem Description

Numerical data that are spatially correlated are often represented in a gridded format. This means that the map over which the numerical data holds, is divided using a raster. Each cell of the raster (or grid) is then assigned a value that is deemed representative for this area. As such, the real world spatial distribution of the modelled value is approximated using a discrete model. Usually, a regular grid with rectangular or square cells is used. Data are often supplied from different sources, and different data are acquired using different technologies. As such, the data are often represented in *incompatible* grids: these are grids that have a different orientation, or different size of grid cells. They are called incompatible, as it is not possible to directly map data from a cell in one grid, to another cell in the other grid. However, this is exactly what needs to be done: scientists want to find correlations between two grids, or assess the influence of one feature onto another feature (e.g. the concentration of air pollutants to which people are exposed). One example is the health impact of airborne pollutants, such as described in [1]. A more complicated example would be judging the benefit of cycling in a city [2]: cycling is good for your health, as it is physical exercise, but cycling in a polluted environment may cause more downsides than benefits.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 366–375, 2014.

There is the exposure to exhaust gasses, but also the changed risk and effects of having an accident, which also needs to be taken into account. Such studies require pollution data, traffic information, accident statistics, traffic patterns and many more. All this information is usually not represented in the same format, and combining the data properly is an issue.

Consider for instance a pollutant that is present over a large area, most likely in different concentrations at different places. The exact distribution might not be fully known (e.g. due to a limited number of measuring points) and is provided as a regular grid with grid cells of e.g. 500m x 500m. Similarly, the population density can also be provided in a gridded format, but its grid cells can have a different size, e.g. 100m x 100m, and even be rotated. Determining which people are exposed to which concentration is a complicated problem, and requires transforming one grid onto the other one. This is illustrated on figure 1: a 4x4 grid has to be remapped onto a 9x9 grid that is slightly angled. If it would be known that the data is related to the black line, the distribution in the 9x9 grid can be better suited, as shown by shaded squares in the examples (a) and (b). Current methods often result in transformations in which the data is more spread out, and moves away from the likely underlying distribution. To overcome this, we present a method that incorporates additional information in order to perform a better transformation of the data.

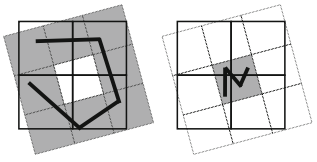



**Fig. 1.** Example of an input grid (2x2, in bold) that needs to be remapped onto a target grid (3x3, dotted line). Different additional data are represented by the thick lines in (a) and (b).

### 1.2 Current Solution Methods

Current solution methods all work on the same principle: the grid that needs to be transformed (this will be called the *input grid*) is analysed and a distribution of the underlying data is assumed. Using this assumed underlying distribution, the data are then remapped to match the grid it will need to be combined with (this will be the *target grid*). A short summary of the most common methods is supplied; for a more detailed overview on different approaches for the map overlay problem, we refer to [3].

The most commonly used is *areal weighting*, [4]. In this approach, the amount of overlap between a gridcell of the input grid and a gridcell of the target grid determines the portion of the associated value of the input gridcell that will be remapped onto the target. Each target gridcell thus gets associated a weighted

sum, where the weights are determined by the amount of overlap with the overlapping input gridcells. This approach assumes that the data in each cell of the input grid are spread uniformly. This assumption however is not always true: in the case of air pollution, the concentration of some pollutants could be linked to linear sources e.g. traffic on roads or could be caused by point sources, which implies that the concentration should be more focussed around the source (taking into account dispersion of the pollutant using existing dispersion models). In figure 1, this means that the data of the 4 input grid cells would be spread out over the 9 target grid cells, not considering the source indicated by the thick line.

A more refined approach to this is *areal smoothing*. In this approach, the data modelled in the input grid is approximated by interpreting the data as a third dimension, and fitting a smooth surface over it. The assumption here is that the data modelled by the input grid are showing a smooth distribution over the whole region of interest. The smooth 3D surface is then resampled using the target raster. This sampling results in the values that will be associated with the cells. While allowing for a smooth transition, the method has the same disadvantage as areal weighting, in that it cannot cope well with local effects such as point or line sources.

## 2 Rulebase Approach

### 2.1 A Different Look at the Problem

The main issue with the problem is that the underlying distribution is not known: the current methods approach the problem by (implicitly) assuming a distribution. Additional knowledge might however be present to help determine a better distribution. An example where one dataset can be improved is when different datasets are fused. In [5], the authors combine different datasets in order to obtain a higher quality dataset. The methodology however is applied on vectorial data that is tagged (e.g. a region tagged as forest, a region tagged as agricultural land, etc). After deriving a common ontology, and after combining the different definitions for regions on the maps, the authors derive a new map that contains the combined information of both.

Generally, when there is a grid representing data, there might be other knowledge that are known to influence the distribution. In the ongoing example of the air pollutant, the type of pollutant and its source can provide information on this. If the particular chemical or particle originates from car exhausts, then the distribution should more or less match the road network (after correction for dispersion). Different pollutants might as such have a different underlying distribution. Such knowledge, makes it possible to make good judgements on the underlying distribution, as shown in [6]. For every grid cell in the target grid, the additional knowledge can be considered. This can be by taking amount over overlap with features of the additional knowledge, the distance to specific items, etc. In [7], a detailed description on how an expert would reason about the redistribution using additional data is presented.

The additional knowledge should only be used to *steer* the distribution, but it should not be followed too strongly: if the additional knowledge is from a different time (e.g. pollution data from 2010, traffic information from 2009), the correlation is weaker. Following that data too strongly might not even be possible and thus would either yield no solution, or a solution that obscures real data. The additional data might also not be the only explanation for the distribution, other sources might be present but unknown. This again is an argument for not too strictly adhering to this information.

### 2.2 Emulating the Intelligent Reasoning

A fuzzy inference system is a system that uses fuzzy sets to represent data and evaluates predicates using simple rules and fuzzy matching [8]. Fuzzy sets are a way of representing uncertain or imprecise information by means of a membership function ([9], [10]). The membership function indicates the possibility or membership of each value. Given an adequate domain, such membership functions can be used to represent e.g. linguistic terms such as *low*: on a domain $[0, 100]$ all values below 10 can the be low (with possibility 1), values above 20 can be considered as *not low* (represented by possibility 0), and values between 10 and 20 have a linearly decreasing membership from 1 to 0. The fuzzy inference system has multiple rules of the form:

`IF x is <linguistic term> THEN y is <linguistic term>`

Here, `x` is a numerical input variable, `y` is the output value, and `<linguistic term>` is a fuzzy set representation for e.g. high, low or other other possible value descriptions. There can be multiple input values, combined using logical operators *and* and *or*. The input variable is matched against the linguistic term, which results in a value in $[0, 1]$ that indicates how well the value matches the term. Based on this, `y` is assigned a linguistic term in its domain. The term is represented by a fuzzy set. There are multiple rules in the rulebase, and `x` can match multiple rules at the same time, resulting in multiple fuzzy sets for `y`. All these results for `y` are aggregated to a single fuzzy set, which is then subsequently defuzzified to yield the crisp result. Several algorithms for defuzzification exist, but for now the most common *center of gravity* will be used.

In [7], we presented how an inference system can be applied to emulate the intelligent reasoning. Key to achieving this is defining the rulebase and the parameters in the rulebase. In order to guarantee that the new distribution still resembles the input distribution, the redistribution of the data happens locally, within a single input cell. The target grid is specified completely independent from the input grid, so first a new grid is computed, the *segment grid*. This grid is made up of all the intersections between input and output cells. Each cell in this grid (for the remainder of this article called *segment*) will only overlap with a single cell from the input grid, and with a single cell from the output grid. Every input cell is completely and exactly covered by a number of segments, as is every output cell. In the algorithm, the segment grid will be used as the new target grid. The problem then becomes a problem of redistributing the data in an input cell over the contained segments. Subsequently, the segments can

be combined differently to form output cells. To facilitate implementation, the additional knowledge is also represented as gridded data. Even if the original knowledge is in a different format (e.g. a road network represented by lines), it is a straight forward operation to convert this to a grid with a small cell size.

### 2.3 Parameters and Range

In order to make the inference system, it is necessary to define parameters. These are values that are considered to provide some correlation with an output: proportional (a high value of the parameter coincides with a high value of the ideal value), or inverse proportional (a higher value of the parameter coincides with a lower value of the ideal value). In [11], several candidates for parameters were proposed. Here, the considered parameters are:

- amount of the auxiliary cell covered by the segment
- amount of the input cell covered by the segment
- amount of the interior of the auxiliary cell covered by the interior of the segment

These parameters were chosen after running several experiments, as they provided the best overall results. Consider the first parameter: "amount of the auxiliary cell covered by the segment". It is intuitive to state that the more of the auxiliary cell is covered by this segment, the higher the value of this segment should be: higher auxiliary value should yield a higher output value. In the rulebase this could be called `aux_overlap`, the value would be used in a rule of the form:

```
IF aux_overlap is low THEN output is low
IF aux_overlap is medium THEN output is medium
IF aux_overlap is high THEN output is high
```

The linguistic terms low, medium and high for `aux_overlap` need to be defined, which means finding adequate limits for the domain of the `aux_overlap` value. When the limits of the domain for the parameter (e.g. `aux_overlap`) are known, a number of equally spaced and shaped triangular fuzzy sets are defined over this domain to define the linguistic terms. The number of triangular fuzzy sets is chosen for each parameter. The more fuzzy sets are defined on the domain, the more rules the rulebase will have; this effectively poses a practical limit. More fuzzy sets should yield more possibilities of distinguishing different values. The main problem now is determining the domain. In our interpretation, the domain is defined by the possible values this particular parameter can have for this segment, thus it varies with each segment. For the relation between segments and the auxiliary grid, there are several possibilities. In the simplest case, the segment covers part of an auxiliary cell. The total value of this auxiliary cell can therefore be considered to be in this segment (e.g. in case of a point source that is the source of the entire value), partly in this segment, or not at all in this segment (e.g., if the value is due to sources outside of this segment - which is

possible as there are other segments overlapping the same cell). The first case results in the maximum possible value. The last case results in the minimum possible value, 0 unless one or more auxiliary cells are fully contained inside the segment, the minimum possible value is then total value of those contained cells. The weighted value is considered as the value of the parameter that is verified, and thus is passed as parameter `x`. The calculation for the range of other parameters is done similarly.

The range for the value of an output cell is initially unknown, but it is limited by the total of its containing segments. For each segment, the output range is from 0 to the value of the overlapping input cell - due to the definition of the segments, there is exactly one. The exact value is obtained using the fuzzy inference system, resulting in a fuzzy set that is then defuzzified. However, the values of all segments that form an input cell should sum up to the value of that input cell. As the fuzzy result for each segment is defuzzified independently, there is no way to guarantee this. Currently, the defuzzified output is considered as a proportion of the total value of all segment: the real range does not matter, so for now the output range for each segment is $[0, 100]$, where 100 is an arbitrarily chosen value. Once all the segment values are calculated and defuzzified, the obtained value is interpreted as a relative amount of the total of all values for segments that overlap this input cell.

### 2.4 Rulebase Construction

The construction of the rulebase at present is fairly rudimentary: after rules that evaluate the first parameter, the addition of each new parameter multiplies the number of rules by the number of linguistic terms for that parameter. It makes every possible combination of each linguistic term for this parameter and the existing rulebase. In the current examples, three parameters, each represented to ten linguistic terms, result in a rulebase that has $10^3$ rules. The range of the output value is expanded with each added parameter: the more parameters say it should be a large value, the larger the value will be. Afterwards, the output range will be rescaled to match the true range. This is a simple way of creating the rulebase, but it results in a very big rulebase in which many rules may never be matched: contradictions between parameters are present.

## 3 Experiments

To test the methodology, different datasets were generated: a geometric test pattern was sampled onto an 12x12 grid (figure 2a) to result the input grid. In the first three test cases, the grid has to be remapped onto a 25x25 grid; the optimal solution (obtained by sampling the geometry onto the grid) is shown on figure 2b, the solution using areal weighting is shown on figure 2c. The fourth test case requires the remapping onto a 25x25 grid that is at a 20° angle, the ideal solution and areal weighting solution are shown on respectively figure 2d and figure 2e.

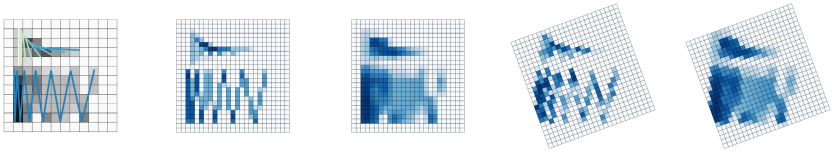



**Fig. 2.** (a) generated input data with grid, (b) ideal solution for target 1, (c) areal weighting for target 1, (d) ideal solution for target 2, (e) areal weighting solution for target 2. Darker shades represent higher associated values; but the scale between different grids does not match. For each grid, black indicates the highest occurring colour in that grid; the lighter the colour, the lower the associated value.

All samples were run using the same three chosen parameters from the previous section; the rulebase system was generated in the same way for all tests, and used ten linguistic variables defined over the domains of each parameter.

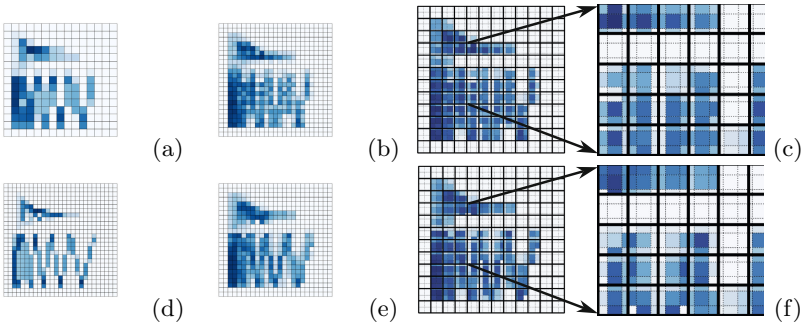



**Fig. 3.** Case 1: low resolution auxiliary data: (a) auxiliary data, (b) result, (c) detail of the remapping of the input data and case 2: High resolution auxiliary data: (d) auxiliary data, (e) result, (f) detail of the remapping of the input data

The developed methodology uses auxiliary data that has to be provided by the user. Experiments were run with different data as auxiliary data, but the auxiliary data was also presented on a grid which was sampled from the same geometries as the input data: this yields *perfect* data, which should provide the best results and allows for the system to be tuned and verified.

In the first test case, $15 \times 15$ auxiliary grid with the same orientation (figure 3a) is used. The result (figure 3b) clearly reveals more detail than areal weighting (figure 2c). The second test case uses a $27 \times 27$ auxiliary grid (figure 3d), and the result shows even more detail (figure 3e). As input and target are the same, it should be compared against the same areal weighting result. The redistribution of the data in the input cells over the segments are shown on figure 3c and figure 3f for both cases: the bold lines show the input grid, the dotted line the output

grid. The segment grid is the irregular grid defined by all these lines. The center part of the segment grid is enlarged for better visibility. On the segment grids, it is clear to see how the value of each input cell is redistributed over its segments. The benefits of the higher resolution auxiliary data are even more visible on this segment grid. Particularly in the second column of the input, the redistribution differs as a result of the different auxiliary data, and the higher values are shifted to the left of those cells. The third test case uses the same target grid, but now

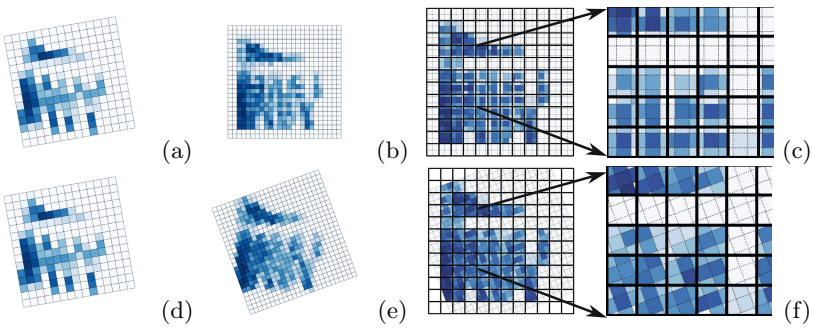



**Fig. 4.** Case 3: low resolution rotated auxiliary data: (a) auxiliary data, (b) result, (c) detail of the remapping of the input data and Case 4: low resolution rotated auxiliary data and rotated target: (a) auxiliary data, (b) result, (c) detail of the remapping of the input data.

employs an auxiliary grid $15 \times 15$ angled $10°$. The fourth test case illustrates the effects if the $25 \times 25$ target grid is angled $20°$. Particularly the distribution of the data inside the cells over the segments in interesting, figures 4c and figure 4f.

To derive a more quantified comparison, consider the absolute difference of the value of a cell in the optimal grid (figure 2b for the first three cases, figure 2d for the fourth case) and the calculated value for the same cell. In table 1, the average and maximum of these differences are shown for both the presented method and for areal weighting. The values for average weighting are the same for the first three cases, as the input and targets are the same. For the results of the presented method, all the values of the second case are lower than those of the first case. This means that the second case has a better result, which is

**Table 1.** Properties of the results of the 4 examples

| | average difference | | maximum difference | |
|---|---|---|---|---|
| | presented | areal weighting | presented | areal weighting |
| case 1 | 3.70 | 3.97 | 41.62 | 33.16 |
| case 2 | 3.11 | 3.97 | 39.74 | 33.16 |
| case 3 | 3.54 | 3.97 | 32.86 | 33.16 |
| case 4 | 7.32 | 7.55 | 81.00 | 82.25 |

also visible on the figures (figure 3b vs. figure 3e). For these cases, the presented method has a lower average difference than areal weighting, but it has a higher maximum average. In simple terms, this means that there are less errors, but larger errors occur. This is consistent with the fact that our methods concentrates the data more, whereas areal weighting tends to smear out the data more over multiple cells: where the line patterns using areal weighting is just visible as a blur, the presented method is able to distinguish more of the pattern. In the case 3 and 4, a low resolution auxiliary grid was used to show that this is enough to contribute. A $15 \times 15$ grid does not add that much information over a $12 \times 12$, but still enough to provide better results. Case 3 shows that the low resolution auxiliary grid at an angle performs slightly worse on average, but better on the maximal difference. In case 4, the values are much higher, as remapping to an angled grid is a more complicated issue. But the presented method still outperforms areal weighting. Compared with the areal weighting approach, the proposed methodology offers better results in remapping the data to the target grid, even when the auxiliary data has a relatively low resolution. The segment grids provide the highest resolution, but unfortunately are irregular. Particularly when input and target are at an angle, the resulting segment grid is not suitable as final representation. The conversion to the target grid is done by adding up all segments that together belong to the same grid cell in the target grid. This effectively lowers the resolution again, which is clearly visible on the figures of the segment grid. However, it results in the desired format. This final step is irreversible: it disconnects the result from the original input grid, and by adding up the values of the segments, the value of an input cell is possibly spread out over a slightly larger region.

## 4 Conclusion

The complexity of the presented method is linear with the number of cells in the segment grid, i.e. the number of cells in the intersection of input and output grid. Consequently, the method scales quite easily. Furthermore, the calculation of each segment can be done independently of other segments, implying they can be computed in parallel. In the above examples, the parameters were manually chosen by us from a large set of parameters ([11]), based on empirical studies on many data. Automatically determining the best possible parameters for a given dataset would improve the applicability. As can be seen on the segmented grids of all examples, but more-so on figure 4, all calculations are constrained within the cells of the input grid.The method tries to localize point sources or line sources at a local level. Mapping the data from the segments to the target grid has the result that data of a segment is spread out over a larger area. As such, it may also give the impression that data are moving out of the original input cells, particularly as the resulting grid is later most likely interpreted as having a uniform distribution within the grid cells. The same applies however to other methods, but as the intermediate results show higher accuracy, perhaps a different aggregation can be considered. In the presented approach, each cell from the input grid is divided in

a number of segments, a possibility distribution for the value of each segment is determined. The value of all segments overlapping an input cell should sum up to the value of the input cell; to achieve this, the defuzzified results were interpreted as relative portions, which required an additional rescaling. The results can be improved by performing a more appropriate defuzzification, and avoiding the rescaling.

In this article, we presented the first experimental results of a novel way to transform gridded data. Unlike current methods, the approach uses additionally known information to estimate an underlying distribution. The presented method uses a fuzzy inference system in order to determine the values of the grid cells in the target. The results are promising, and further research in both refining the system and testing it with real world data are foreseen.

**Acknowledgement.** The study was conducted within the 7FP Marie Curie Actions IRSES project No. 247645, acronym GESAPU.

## References

1. Tainio, M., Sofiev, M., Hujo, M., Tuomisto, J.T., Loh, M., Jantunen, M.J.: Evaluation of the european population intake fractions for european and finnish anthropogenic primary fine particulate matter emissions. Atmospheric Environment 43(19), 3052–3059 (2009)
2. Woodcock, J., Tainio, M., Cheshire, J., O'Brien, O., Goodman, A.: Health effects of the london bicycle sharing system: health impact modelling study. British Medical Journal 348(2) (2014)
3. Gotway, C.A., Young, L.J.: Combining incompatible spatial data. Journal of the American Statistical Association 97(458), 632–648 (2002)
4. Flowerdew, R., Green, M.: Areal interpolation and types of data. In: Foterhingham, S., Rogerson, P. (eds.) Spatial analysis and GIS, pp. 141–152. Taylor & Francis (1994)
5. Duckham, M., Worboys, M.: An algebraic approach to automated information fusion. International Journal of Geographic Information Systems 19, 537–558 (2005)
6. Verstraete, J.: Using a fuzzy inference system for the map overlay problem. In: 3rd International Workshop on Uncertainty in Greenhouse Gas Inventories, pp. 289–298 (2010)
7. Verstraete, J.: Solving the map overlay problem with a fuzzy approach. Climatic Change, 1–14 (2014)
8. Mendel, J.M.: Uncertain rule-based fuzzy logic systems, Introduction and new directions. Prentice Hall (2001)
9. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
10. Dubois, D., Prade, H.: The three semantics of fuzzy sets. Fuzzy Sets and Systems 90, 141–150 (1999)
11. Verstraete, J.: Parameters to use a fuzzy rulebase approach to remap gridded spatial data. In: Proceedings of the 2013 Joint IFSA World Congress NAFIPS Annual Meeting (IFSA/NAFIPS), pp. 1519–1524 (2013)

# Fast and Incremental Computation for the Erosion Score

Gilles Moyse and Marie-Jeanne Lesot

Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France
CNRS, UMR 7606, LIP6, F-75005, Paris, France
firstname.lastname@lip6.fr

**Abstract** The erosion score is a Mathematical Morphology tool used primarily to detect periodicity in data. In this paper, three new computation methods are proposed, to decrease its computational cost and to allow to process data streams, in an incremental variant. Experimental results show the significant computation time decrease, especially for the efficient levelwise incremental approach which is able to process a one million point data stream in 1.5s.

**Keywords:** Mathematical Morphology, Erosion Score, Incremental.

## 1 Introduction

Mathematical Morphology (MM) defines a set of techniques for the analysis of spatial structures, and is widely used in image processing, understanding, segmentation or compression [10,13]. Functional MM applies its principles to function values and has been used for several types of data processing tasks, such as such as signal sieving [1,14], signal pre-processing [17], text categorisation [4], fuzzy classes identification [5] or gradual rule extraction [11].

This paper focuses on the erosion score operator, that has been applied to efficiently detect periodicity in time series, interpreted as functions associating $x_t$ at each time $t$ [9]. The erosion score is based on the fundamental MM erosion operator and is more precisely defined as the sum of successive erosions until total erosion.

This paper considers the implementation issue for this operation and proposes three computation methods both to decrease the computational cost and to allow to process data incrementally: the proposed levelwise approach is based on the identification of levels in the data, to reduce the number of steps required to compute the erosion score for all data points. On the other hand, the proposed incremental extensions of both the basic and the levelwise computation methods make it possible to progressively update erosion scores when new data points become available, so as to process data streams.

The paper is organised as follows: Section 2 recalls the general MM principles and the erosion score as well as existing works related to optimised methods to compute MM operations. The following sections then respectively introduce the Levelwise, Incremental and Levelwise Incremental approaches. Lastly, Section 6 presents the experiments carried out to compare the proposed approaches.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 376–385, 2014.

## 2 Context and Motivations

**Erosion Score Operation.** Mathematical Morphology relies on two basic operators, erosion and dilation, combined in various ways to define more complex composed operators (see [13] for a detailed presentation). In functional MM, given a function $f : E \to F$ and a *structuring element* $B$ defined as a subset of $E$ of a known shape, e.g. an interval centred at the origin, *erosion* is the function $\epsilon_B(f) : E \to F$ defined as $[\epsilon_B(f)](x) = \inf_{b \in B} f(x + b)$. Dilation is defined in a similar way, using the sup operator. These two basic operations can be used repeatedly or alternatively, leading to different types of composed operators, such as opening, closing or alternated filters [13].

The operator proposed in [9] to detect periodicity relies on the computation and aggregation of successive erosions. Given a time series $X$ containing $n$ values $\{x_1, \ldots x_n\}$ in $[0, 1]$ obtained at regular time intervals and the structuring element $B = (-1, 0, 1)$, the previous erosion definition induces the following recursive form for the $j^{th}$ erosion of the $i^{th}$ value: $x_i^j = \min(x_{i-1}^{j-1}, x_i^{j-1}, x_{i+1}^{j-1})$, denoting by convention $x_i^0 = x_i$, $x_0 = x_{n+1} = +\infty$. The iterativity property of this erosion yields $x_i^j = \min(x_{i-j}, \ldots, x_{i+j})$.

The erosion score is defined for a minimum zero $X$ satisfying the property $\exists i \in \{1...n\}$ such that $x_i = 0$ as the sum of the successive erosions until total erosion: denoting $z_i$ the number of erosion steps needed for $x_i$ to be totally eroded, i.e. the smallest erosion step $j$ such that $x_i^j = 0$

$$es_i = \sum_{j=0}^{z_i} x_i^j \tag{1}$$

Other methods based on successive erosions mostly aim at being used in 2D contexts, as the erosion curve [6] or the ultimate erosion [3].

**Implementation Optimisation.** Efficient implementation methods have been proposed for many MM operations, to make it possible to compute them faster [16] or to allow them to process data incrementally [2].

In particular, various optimisations to quickly compute successive erosions have been proposed (see [7] for a recent state of the art): for instance they reduce the number of redundant operations when using various structuring elements [12] or propose a two pass optimisation to ensure a constant time computation of an erosion for any structuring element. However, such methods are not relevant for the erosion score computation where a single and simple structuring element is considered.

Another category of methods rely on the identification of specific values in the dataset, called anchors [15] or obtained after filtering out useless values to reduce the computation time [2]. The methods proposed in the paper belong to this category but they use another definition for the values of interest, as the latter are used to compute a different score.

## 3 Levelwise Method

The "naive" implementation of the erosion score consists in computing the successive erosions of the dataset and summing the obtained values until all eroded values equal 0. Since $x_i$ is eroded in $z_i$ iterations by definition of $z_i$ and the whole dataset is processed at each iteration, its complexity is $O((\max_i z_i) \times n)$.

In this section, a levelwise approach is proposed: it does not process the whole dataset at each iteration but reduces the number of iterations for each data point individually. It is based on the identification of key erosion steps, i.e. a subset of the initial dataset sufficient to compute the erosion scores.

### 3.1 Notations

When computing the successive eroded values $x_i^j$, it can be observed that some of them are equal to the previous one. Identifying only the key ones, defined as the $x_i^j$ different from the previous one, allows to compute the erosion score by adding the distinct values multiplied by their number of repetitions.

Formally, the key erosion steps are such that $x_i^j \neq x_i^{j-1}$, i.e. $x_i^j < x_i^{j-1}$ due to their definition. Let us denote $J_i^< = \{j \in \{1, \dots, n\} | x_i^j < x_i^{j-1}\}$, $\omega_i$ its size and $D_i$ the ordered set of its values, sorted in ascending order, to which 0 is added as $d_{i0}$: $D_i = \{d_{i0}, \dots, d_{i\omega_i}\}$ is an ordered subset of $\{0, \dots, n\}$ where only the key steps are kept, in that the erosion score can be computed knowing them only. It can be noted that the maximal value of $D_i$ is $d_{i,\omega_i} = z_i$. It holds that $\forall l \in \{0, \dots, \omega_i - 1\}$, $d_{il} < d_{il+1}$ and $x_i = x_i^{d_{i0}} = x_i^{d_{i0}+1} = \dots = x_i^{d_{i1}-1} > x_i^{d_{i1}} = x_i^{d_{i1}+1} = \dots = x_i^{d_{i2}-1} > x_i^{d_{i2}}$ and so on until $x_i^{d_{i,\omega_i}-1} > x_i^{d_{i,\omega_i}} = 0$.

We also introduce the notations $\chi_{il} = x_i^{d_{il}}$ and $\lambda_{il}$ its index such that $\chi_{il} = x_{\lambda_{il}}$. $d_{il}$ can be seen as the number of points between $x_i$ and its $l^{th}$ key value, $\chi_{il}$ is its value and $\lambda_{il}$ its index. An illustrative example is given in Fig. 1.

### 3.2 Levelwise Computation of the Erosion Score

For any data point $x_i$, $D_i$ then contains the key erosions, so the erosion score can be computed knowing these values only, as stated in the following theorem:

**Theorem 1.** *Levelwise computation of* $es_i$

$$es_i = \sum_{l=0}^{\omega_i - 1} (d_{i,l+1} - d_{il})\, \chi_{il} = \sum_{l=0}^{\omega_i - 1} (|\lambda_{i,l+1} - i| - |\lambda_{il} - i|) x_{\lambda_{il}}$$

*Proof.* The demonstration directly follows from the definitions of $\chi$, $\lambda$ and $d$ developing the definition given in Eq. (1)

$$es_i = \underbrace{x_i^0 + \dots + x_i^{d_{i1}-1}}_{(d_{i1}-d_{i0})\chi_{i0}} + \underbrace{x_i^{d_{i1}} + \dots + x_i^{d_{i2}-1}}_{(d_{i2}-d_{i1})\chi_{i1}} + \dots + \underbrace{x_i^{d_{i,\omega_i-1}} + \dots + x_i^{d_{i,\omega_i}-1}}_{(d_{i\omega_i}-d_{i,\omega_i-1})\chi_{i,\omega_i-1}} + \underbrace{x_i^{d_{i,\omega_i}}}_{0}$$

$$= \sum_{l=0}^{\omega_i - 1} (d_{i,l+1} - d_{il})\, \chi_{il}$$

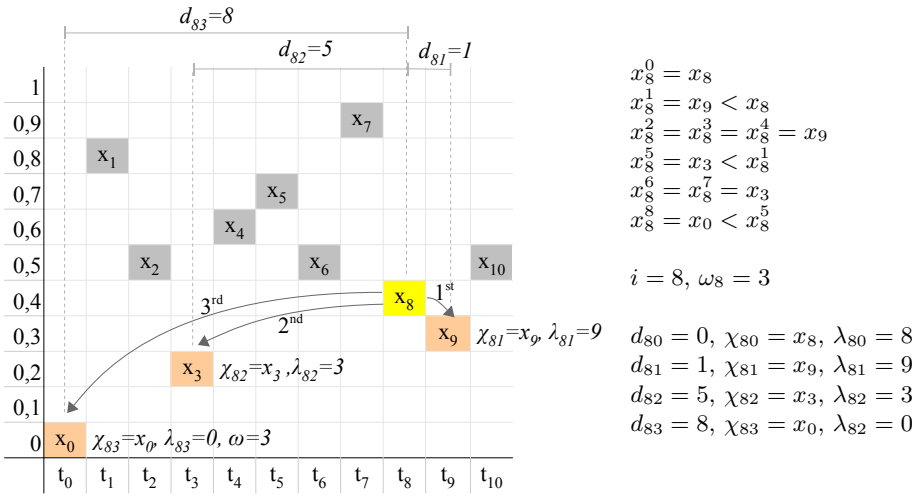



**Fig. 1.** Example and computation of $D_8 = \{d_{80}, d_{81}, d_{82}, d_{83}\}$, $\chi_{8l}$ and $\lambda_{8l}$

The expression based on $\lambda_{il}$ is obtained from the fact that $d_{il} = |\lambda_{il} - i|$.

Based on this theorem, the levelwise implementation is computed from the $\lambda_{il}$ values only. They are computed by searching $x_i$ key erosions until a zero is reached, i.e. for $j$ from 1 to $z_i$, so in $z_i$ iterations. Since the erosion score is computed at the same time, the complexity of this method is $O\left(\sum z_i\right)$ which is lower than the naive one presented previously.

## 4 Incremental Method

Incremental approaches aim at processing the data successively, updating intermediate results instead of considering each data point independently. They can decrease the computational time; moreover they make it possible to process data streams, where the whole dataset is not available at once but data points are obtained progressively.

In this section and the following, we propose incremental methods to compute the erosion score, respectively based on the basic and the levelwise approaches.

### 4.1 Notations

In the context of data streams, $x_{n+1}$ denotes the latest data point received at time $t+1$. We denote $x_i(t) = x_i$ if $i \in \{1, \ldots, n\}$ and $x_i(t) = +\infty$ otherwise. For simplicity sake, the notation $x_i$ is preferred to $x_i\,(t)$ when no ambiguity arises. The value $x_0 = 0$ is added in order to ensure the minimum zero property and thus that the algorithm terminates. At time $t$, the $j^{th}$ erosion of $x_i$ is denoted $x_i^j\,(t)$ and its erosion score $es_i\,(t)$.

The objective of incremental methods is to compute the new erosion scores $es_i\,(t+1)$ from the existing erosion scores $es_i\,(t)$ and the new value $x_{n+1}$.

### 4.2 Update Equations for Eroded Values and Erosion Score

The theorem below establishes the update equation that gives the new eroded values for any data point when a new data point $x_{n+1}$ is collected.

**Theorem 2.** *Update equations for the successive eroded values*
*Denoting* $q = \max\{k \in \{1,\dots,n\} \,|\, x_k \leq x_{n+1}\}$ *and* $m = (n+1+q)/2$,

$$x_i^j(t+1) = \begin{cases} x_i^j(t) & \text{if } i \leq m \\ x_i^j(t) & \text{if } i > m \text{ and } j < n+1-i \\ x_{n+1} & \text{if } i > m \text{ and } n+1-i \leq j < i-q \\ x_q^{j-(i-q)} & \text{if } i > m \text{ and } j \geq i-q \end{cases}$$

*Proof.* $q$ is the index of the latest data point less than or equal to the new point $x_{n+1}$ and $m$ the middle of the index interval between $q$ and $n+1$.

The proof consists in studying, for any $i$ and $j$ whether $x_{n+1}$ and/or $x_q$ are involved in the computation of $x_i^j(t+1)$. Since $x_i^j = \min(x_{i-j},\dots,x_{i+j})$, this is equivalent to checking whether $n+1$ and/or $q$ belongs to $\{i-j,\dots,i+j\}$.

If $x_{n+1}$ is not involved, then $x_i^j(t+1) = x_i^j(t)$. This is the case if the data point is closer to $x_q$ than to $x_{n+1}$, so when $i \leq m$, since $x_q \leq x_{n+1}$ by definition. If $i > m$ and $j < n+1-i$, $x_{n+1}$ is not involved too, since $n+1 \notin \{i-j,...i+j\}$.

If $i > m$ and $n+1-i \leq j < i-q$, then $x_{n+1}$ is involved but not $x_q$, so $x_i^j(t+1) = x_{n+1}$. Indeed, for all $l \in \{q+1,\dots,n+1\}$, $x_l \geq x_{n+1}$ and the minimal data value on all index intervals included in $\{q+1,\dots,n+1\}$ is $x_{n+1}$.

Finally, if $i > m$ and $j \geq i-q$, then both $x_{n+1}$ and $x_q$ are involved, so $x_i^j(t+1) \leq x_q \leq x_{n+1}$, by definition of $x_q$. Therefore:

$$\begin{aligned} x_i^j(t+1) &= \min(x_{i-j},\dots,x_q,\dots,x_{n+1},\dots,x_{i+j}) = \min(x_{i-j},\dots,x_q) \\ &= \min(x_{q-(j-i+q)},\dots,x_q,\dots,x_{q+(j-i+q)}) = x_q^{j-i+q} \end{aligned}$$

These update equations lead to the update equations for the erosion score:

**Theorem 3.** *Computation of* $es\,(t+1)$
*Denoting* $q = \max\{k \in \{1,\dots,n\} \,|\, x_k \leq x_{n+1}\}$ *and* $m = (n+1+q)/2$,

$$es_i\,(t+1) = \begin{cases} es_i\,(t) & \text{if } i \leq m \\ es_q\,(t) + 2\,(i-m)\,x_{n+1} + \sum_{j=0}^{n-i} x_i^j\,(t) & \text{otherwise} \end{cases}$$

*Proof.* The theorem is a direct consequence of Theorem 2: if $i \leq m$, the successive erosions are not modified so neither is their sum. If $i > m$, the following decomposition of the erosion score proves the theorem:

$$es_i\,(t+1) = \underbrace{\sum_{j=0}^{n-i} x_i^j\,(t+1)}_{=\sum_{j=0}^{n-i} x_i^j(t)} + \underbrace{\sum_{j=n+1-i}^{i-q-1} x_i^j\,(t+1)}_{=2(i-m)x_{n+1}} + \underbrace{\sum_{j=i-q}^{+\infty} x_i^j\,(t+1)}_{=es_q(t)}$$

It is easily proven that the complexity of the incremental method is $O\left(\phi^2\right)$, where $\phi = n - q$, i.e. the distance between the new point and the latest lower data point.

## 5 Incremental Levelwise Method

This section proposes an alternative incremental method, based on the levelwise expression of the erosion score stated in Theorem 1 and on incremental update equations for the $\lambda$ indices.

**Theorem 4.** *Incremental computation of* $\lambda_{il}$

*Denoting* $q = \max\{k \in \{1,\ldots,n\} \text{ st } x_k < x_{n+1}\}$, $m = (n+1+q)/2$ *and* $k_i$ *defined for* $i > m$ *such that* $d_{i,k_i-1}(t) < n+1-i \leq d_{ik_i}(t)$ *and* $k_{n+1} = 0$

$$\forall i, \forall l,\ \lambda_{il}(t+1) = \begin{cases} \lambda_{il}(t) & \text{if } i \leq m \\ \lambda_{il}(t) & \text{if } i > m \text{ and } l < k_i \\ n+1 & \text{if } i > m \text{ and } l = k_i \\ \lambda_{q,l-k_i-1}(t) & \text{if } i > m \text{ and } l > k_i \end{cases}$$

$$\forall i,\ \omega_i(t+1) = \begin{cases} \omega_i(t) & \text{if } i \leq m \\ k_i + \omega_q(t) & \text{if } i > m \end{cases}$$

*Proof.* The incremental expression stated in Theorem 2 allows the update of $x_i^j$ for $j = 0...z_i$. Since $D_i$ is a subset of $0...z_i$ containing only the key erosions, this proof is based on the one presented for the incremental method. $k_i$ is introduced to represent the index in $D_i$ denoting the first erosion involving $x_{n+1}$.

As in the incremental case, if $i \leq m$, or if $i > m$ and $x_{n+1}$ is not involved, i.e. $l < k_i$, then the successive erosions are unchanged, so $\lambda_{il}(t) = \lambda_{il}(t+1)$.

If $l = k_i$ then the eroded value is $x_{n+1}$, so its index $\lambda_{i,k_i}$ is $n+1$.

Finally, as proved for the incremental method, the next erosions, following the one equal to $x_{n+1}$, are the erosions of $x_q$. In the levelwise context, it implies that the key erosions following the one equal to $x_{n+1}$ are also the key erosions of $x_q$ for $l = 0...\omega_q$, so for $i > m$ and $l > k_i$, $\lambda_{il}(t+1) = \lambda_{q,l-k_i-1}(t)$.

Regarding $\omega_i$ the number of elements in $D_i$, when the eroded values are unchanged, i.e. $i \leq m$, the number of key erosions is unchanged too, and $\omega_i(t+1) = \omega_i(t)$. If $i > m$, then the key erosions are those from 0 to the first implying $x_{n+1}$ whose index is $k_i$, plus the $\omega_q$ key erosions of $x_q$, yielding $\omega_i(t+1) = k_i + \omega_q(t)$.

This theorem together with the levelwise expression of the erosion score stated in Theorem 1 lead to the following levelwise incremental expression of the erosion score:

**Theorem 5.** *Incremental levelwise computation of $es_i$*

*Denoting $m$, $q$ and $k_i$ as defined in Theorem 4 and $p_i$ defined for $i \geq m$ such that $\lambda_{ip_i}(t) = q$:*

$$\forall i,\ es_i(t+1) = \begin{cases} es_i(t) & \text{if } i \leq m \\ \chi_{i,k_i-1}(t)(n+1-i-d_{ik_i}(t)) & \text{if } m < i < n+1 \\ \quad -\sum_{j=k_i}^{p_i-1} \chi_{ij}(t)(d_{i,j+1}(t) - d_{ij}(t)) & \\ \quad +2x_{n+1}(i-m) + es_i(t) & \\ 2x_{n+1}(n+1-m) + es_q(t) & \text{if } i = n+1 \end{cases}$$

The variables $d$ and $\chi$ are used to improve readability, but $es_i(t+1)$ can be computed with $\lambda$ only since $d_{il} = |\lambda_{il} - i|$ and $\chi_{il} = x_{\lambda_{il}}$.

The proof is omitted because of space constraints. It follows from the decomposition of the sum given in Theorem 1 into 4 terms, corresponding to $l$ lower than $k_i - 2$, equal to $k_i - 1$, between $k_i$ and $p_i$, and greater than $p_i$. In each term, the $\lambda_{il}(t+1)$ values are replaced by their expression given by Theorem 4.

The implementation then consists in using a $\Lambda$ matrix storing all $\lambda_{il}$ values. When a new data point $x_{n+1}$ is processed, its predecessor $x_q$ is first identified. Then for each row whose index is greater than $m$, $\Lambda$ is updated by computing $k_i$, inserting $n+1$ as the $k_i^{th}$ value in the list, and copying the values from $(\lambda_{ql})_l$ at index $k_i+1$.

## 6 Experiments

### 6.1 Experimental Protocol

The 4 methods, namely "naive", "incremental", "levelwise", and "incremental levelwise", are compared over artificial datasets generated as noisy repetitions of identical blocks of different shapes (triangle, rectangle, sine, wave). Noise applies either on the size of the individual blocks, randomly enlarging or shrinking them, or on the data points, randomly adding or subtracting small values (see [8] for a more detailed presentation of the data generation process).

Each method is then applied to compute the erosion score of each data point in each dataset. For a given dataset, all methods return the same result. The data points are read one by one for the incremental approaches so as to emulate a stream.

For each method, the computational time as well as the memory consumption are measured; their average and standard deviation are computed over all the datasets randomly generated according to the protocol sketched above. Furthermore, the average value of $\omega_i$ is recorded for the levelwise methods.

The implementation is done in VB.NET and the experiments are run on a Windows® virtual machine started with 4 CPUs and 4 Go of memory on a physical machine with an Intel i7® CPU and 16 Go of memory.

### 6.2 Computational Time Comparison

Figure 2 shows the computational time for three levels of dataset sizes: the top graph compares all 4 methods for datasets containing less than 10,000 data points. It shows that the incremental methods outperform the non incremental ones. The naive implementation is significantly slower and less robust as the high standard deviation shows. Furthermore, the incremental methods run much faster in a real situation since the arrival of a new data implies only one computation whereas the non incremental ones have to run the computation anew over the whole dataset.

In order to differentiate more precisely the incremental methods, larger datasets are used, ranging from 10,000 to 100,000 data points as showed on the middle graph of Fig. 2. In this second round of experiments, the incremental levelwise method appears significantly faster than the incremental one. Moreover, the large values of standard deviation for the latter indicate a lack of robustness. This is due to the sensitivity of the position of zero values within the dataset for the methods not based on the levelwise approach. Indeed, as underlined in the complexity analysis of the 2 methods (see Section 2), if the largest $z_i$ in a dataset is increased by only 1, another full scan of the dataset is needed with the naive method. With the levelwise approach on the contrary, one more iteration is needed only for the concerned data point

Finally, the largest datasets (bottom graph on Fig. 2) show that the incremental levelwise approach for erosion score computation is robust and can handle efficiently and in a very reasonable amount of time a large dataset on the order of one million data points. Moreover, since it is incremental, it can handle a million new data from one stream in 1.5 seconds, or equivalently handle 1 new data over a million streams in the same time.

### 6.3 Memory Consumption

In terms of memory, the non levelwise methods, whether incremental or not, are not based on specific data structures and thus do not need more memory than the space needed by the dataset and the resulting erosion scores: denoting $n$ the number of data points, the memory consumption in this case is $2n$.

In the levelwise methods, the $\Lambda$ matrix is stored, entailing an additional memory usage: it is implemented as a list of $n$ lists each of them containing $\omega_i$ values. Its memory requirements is then $\sum \omega_i$ or equivalently $n \times avg(\omega_i)$. Over all carried out experiments, the average $\omega_i$ is 30, the minimum 2 and maximum 129, thus the required storing capacities remain reasonable.

Hence, the levelwise methods are more greedy in terms of memory than the non levelwise ones. Nonetheless, this can be mitigated simply for the incremental levelwise since when a zero value is reached, all previous values become useless in the erosion score, so the $\lambda_i$ before the 0 value can be removed from $\Lambda$.

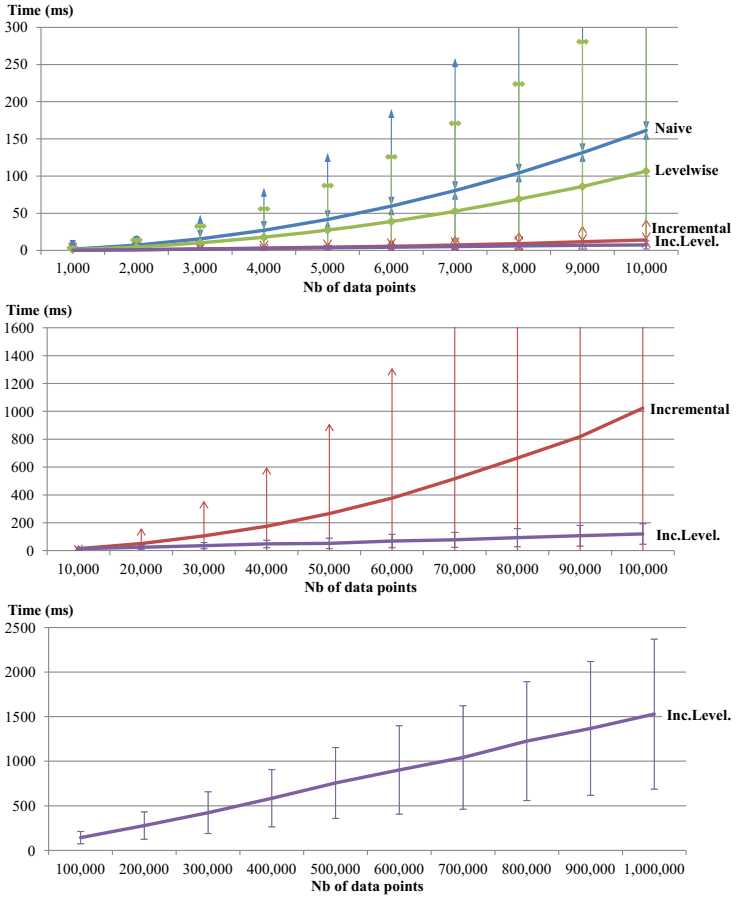



**Fig. 2.** Computational time for (top) small datasets and all 4 methods, (middle) medium datasets and incremental methods, (bottom) large datasets and the incremental levelwise method

## 7 Conclusion and Future Works

This paper proposed 3 variants to compute the erosion score based on one hand on a levelwise computation principle and on the other hand on update equations to progressively adapt to new incoming data points. Such incremental approaches make it possible to process data streams where data points are not available simultaneously. Experimental studies show the relevance of these variants and in particular the performance of the levelwise incremental approach, in terms of time consumption at the expense of a reasonable increase of memory storage.

Future works aim at integrating the efficient levelwise incremental method to the periodicity detection task, to identify periodicity in large time series. Other perspective include the use of this approach to other time series pre-processing tasks where the series structure must be identified.

## References


1. Bangham, A., Marshall, S.: Image and signal processing with mathematical morphology. Electronics & Communication Engineering Journal 10(3), 117–128 (1998)
2. Dokládal, P., Dokládalová, E.: Computationally efficient, one-pass algorithm for morphological filters. Journal of Visual Communication and Image Representation 22(5), 411–420 (2011)
3. Lantuejoul, C., Maisonneuve, F.: Geodesic methods in quantitative image analysis. Pattern Recognition 17(2), 177–187 (1984)
4. Lefèvre, S., Claveau, V.: Topic segmentation: application of mathematical morphology to textual data. In: Soille, P., Pesaresi, M., Ouzounis, G.K. (eds.) ISMM 2011. LNCS, vol. 6671, pp. 472–481. Springer, Heidelberg (2011)
5. Marsala, C., Bouchon-Meunier, B.: Choice of a method for the construction of fuzzy decision trees. In: Proc. of FUZZ-IEEE 2003, vol. 1, pp. 584–589 (2003)
6. Mattioli, J., Schmitt, M.: Inverse problems for granulometries by erosion. Journal of Mathematical Imaging and Vision 2(2-3), 217–232 (1992)
7. Morard, V., Dokládal, P., Decencière, E.: One-Dimensional Openings, Granulometries and Component Trees in Per Pixel. IEEE Journal of Selected Topics in Signal Processing 6(7), 840–848 (2012)
8. Moyse, G., Lesot, M.-J., Bouchon-Meunier, B.: Mathematical morphology tools to evaluate periodic linguistic summaries. In: Larsen, H.L., Martin-Bautista, M.J., Vila, M.A., Andreasen, T., Christiansen, H. (eds.) FQAS 2013. LNCS, vol. 8132, pp. 257–268. Springer, Heidelberg (2013)
9. Moyse, G., Lesot, M.-J., Bouchon-Meunier, B.: Linguistic summaries for periodicity detection based on mathematical morphology. In: Proc. of IEEE SSCI FOCI 2013, pp. 106–113 (2013)
10. Najman, L., Talbot, H.: Mathematical Morphology. John Wiley & Sons (2013)
11. Oudni, A., Lesot, M.-J., Rifqi, M.: Characterisation of gradual itemsets based on mathematical morphology tools. In: Proc. of EUSFLAT 2013, pp. 826–833 (2013)
12. Pecht, J.: Speeding-up successive Minkowski operations with bit-plane computers. Pattern Recognition Letters 3(2), 113–117 (1985)
13. Serra, J.: Introduction to mathematical morphology. Computer Vision, Graphics, and Image Processing 35(3), 283–305 (1986)
14. Sun, Y., Chan, K.L., Krishnan, S.M.: Characteristic wave detection in ECG signal using morphological transform. BMC Cardiovascular Disorders 5(1), 28 (2005)
15. Van Droogenbroeck, M., Buckley, M.: Morphological erosions and openings: fast algorithms based on anchors. Journal of Mathematical Imaging and Vision 22(2-3), 121–142 (2005)
16. Vincent, L.: Fast grayscale granulometry algorithms. In: ISMM, vol. 2, pp. 265–272 (1994)
17. Wang, X., Tang, H., Zhao, X.: Noisy Speech Pitch Detection Based on Mathematical Morphology and Weighted MACF. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) SINOBIOMETRICS 2004. LNCS, vol. 3338, pp. 594–601. Springer, Heidelberg (2004)

# Complexity of Rule Sets Induced from Incomplete Data Sets Using Global Probabilistic Approximations

Patrick G. Clark[1] and Jerzy W. Grzymala-Busse[1,2]

[1] Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045, USA
[2] Department of Expert Systems and Artificial Intelligence, University of Information Technology and Management, 35-225 Rzeszow, Poland
patrick.g.clark@gmail.com, jerzy@ku.edu

**Abstract.** We consider incomplete data sets using two interpretations of missing attribute values: lost values and "do not care" conditions. Additionally, in our data mining experiments we use global probabilistic approximations (singleton, subset and concept). The results of validation of such data, using global probabilistic approximations, were published recently. A novelty of this paper is research on the complexity of corresponding rule sets, in terms of the number of rules and number of rule conditions. Our main result is that the simplest rule sets are induced from data sets in which missing attribute values are interpreted as "do not care" conditions where rule sets are induced using subset probabilistic approximations.

## 1 Introduction

Probabilistic approximations, for complete data sets and based on an equivalence relation, were studied for many years [14–19]. Incomplete data sets may be analyzed using global approximations such as singleton, subset and concept [5–7]. Probabilistic approximations, for incomplete data sets and based on arbitrary binary relations, were introduced in [8], while first experimental results using probabilistic approximations were published in [1].

In this paper incomplete data sets are characterized by missing attribute values. We will use two interpretations of a missing attribute value: *lost values* and *"do not care" conditions*. Lost values indicate the original value was erased or never obtained, and as a result we should use only existing, specified attribute values for rule induction. "Do not care" conditions identify data that may be replaced by any specified attribute value, typically someone refused to answer a question.

A probabilistic approximation is defined using a probability $\alpha$. If $\alpha$ is equal to one, the probabilistic approximation is equal to the lower approximation; if $\alpha$ is a sufficiently small, positive number, the probabilistic approximation is equal to the upper approximation. Both lower and upper approximations are fundamental ideas of rough set theory.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 386–395, 2014.

The main objective of this paper is research on the complexity of rule sets, in terms of the number of rules and number of rule conditions, induced from data sets with lost values and "do not care" conditions, while rule sets are induced using three global approximations: singleton, subset and concept. These approximations and their relationship to probabilistic approximations are defined in section 3. Our main result is that the simplest rule sets are induced from data sets in which missing attribute values are interpreted as "do not care" conditions where rule sets are induced using subset probabilistic approximations.

## 2 Attribute-Value Pair Blocks

We assume that the input data sets are presented in the form of a *decision table*. Example of decision tables are shown in Tables 1 and 2. Rows of the decision table represent *cases*, while columns are labeled by *variables*. The set of all cases will be denoted by $U$. In Tables 1 and 2, $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$. Some variables are called *attributes* while one selected variable is called a *decision* and is denoted by $d$. The set of all attributes will be denoted by $A$. In Tables 1 and 2, $A = \{Wind, Humidity, Temperature\}$ and $d = Trip$.

An important tool to analyze data sets is a *block of an attribute-value pair*. Let $(a, v)$ be an attribute-value pair. For *complete* decision tables, i.e., decision tables in which every attribute value is specified, a block of $(a, v)$, denoted by $[(a, v)]$, is the set of all cases $x$ for which $a(x) = v$, where $a(x)$ denotes the value of the attribute $a$ for the case $x$. For incomplete decision tables the definition of a block of an attribute-value pair is modified [5–7].

– If for an attribute $a$ there exists a case $x$ such that $a(x) = ?$, i.e., the corresponding value is lost, then the case $x$ should not be included in any blocks $[(a, v)]$ for all values $v$ of attribute $a$,

**Table 1.** An incomplete decision table with lost values

| | Attributes | | | Decision |
|---|---|---|---|---|
| Case | Wind | Humidity | Temperature | Trip |
| 1 | low | ? | high | yes |
| 2 | ? | ? | high | yes |
| 3 | high | high | ? | yes |
| 4 | ? | low | low | yes |
| 5 | ? | high | low | no |
| 6 | low | ? | low | no |
| 7 | high | high | high | no |
| 8 | high | high | ? | no |

– If for an attribute $a$ there exists a case $x$ such that the corresponding value is a "do not care" condition, i.e., $a(x) = *$, then the case $x$ should be included in blocks $[(a, v)]$ for all specified values $v$ of attribute $a$.

**Table 2.** An incomplete decision table with "do not care" conditions

| | Attributes | | | Decision |
|---|---|---|---|---|
| Case | Wind | Humidity | Temperature | Trip |
| 1 | low | * | high | yes |
| 2 | * | * | high | yes |
| 3 | high | high | * | yes |
| 4 | * | low | low | yes |
| 5 | * | high | low | no |
| 6 | low | * | low | no |
| 7 | high | high | high | no |
| 8 | high | high | * | no |

**Table 3.** Blocks $[(a, v)]$ of attribute value pairs $(a, v)$

| | Lost values | "Do not care" conditions |
|---|---|---|
| [(Wind, low)] | {1, 6} | {1, 2, 4, 5, 6} |
| [(Wind, high)] | {3, 7, 8} | {2, 3, 4, 5, 7, 8} |
| [(Humidity, low)] | {4} | {1, 2, 4, 6} |
| [(Humidity, high)] | {3, 5, 7, 8} | {1, 2, 3, 5, 6, 7, 8} |
| [(Temperature, low)] | {4, 5, 6} | {3, 4, 5, 6, 8} |
| [(Temperature, high)] | {1, 2, 7} | {1, 2, 3, 7, 8} |

A block of a decision-value pair is called a *concept*. In Tables 1 and 2, the concepts are [(Trip, yes)] = {1, 2, 3, 4} and [(Trip, no)] = {5, 6, 7, 8}. Table 3 presents the attribute-value blocks computed for Table 1 (lost values) and Table 2 ("do not care" conditions).

Let $B$ be a subset of the set $A$ of all attributes. For a case $x \in U$ the *characteristic set* $K_B(x)$ is defined as the intersection of the sets $K(x, a)$, for all $a \in B$, where the set $K(x, a)$ is defined in the following way:

– If $a(x)$ is specified, then $K(x, a)$ is the block $[(a, a(x))]$ of attribute $a$ and its value $a(x)$,

– If $a(x) =?$ or $a(x) = *$ then the set $K(x, a) = U$.

**Table 4.** Characteristic sets for the entire attribute set $A$

| Characteristic sets | | |
|---|---|---|
| Case | Lost values | "Do not care" conditions |
| 1 | $\{1\}$ | $\{1, 2\}$ |
| 2 | $\{1, 2, 7\}$ | $\{1, 2, 3, 7, 8\}$ |
| 3 | $\{3, 7, 8\}$ | $\{2, 3, 5, 7, 8\}$ |
| 4 | $\{4\}$ | $\{4, 6\}$ |
| 5 | $\{5\}$ | $\{3, 5, 6, 8\}$ |
| 6 | $\{6\}$ | $\{4, 5, 6\}$ |
| 7 | $\{7\}$ | $\{2, 3, 7, 8\}$ |
| 8 | $\{3, 7, 8\}$ | $\{2, 3, 5, 7, 8\}$ |

For example, the characteristic set for case 1 from Table 1 is

$$K_A(1) = [(Wind, low)] \cap U \cap [(Temperature, high)]$$
$$= \{1, 6\} \cap \{1, 2, 3, 4, 5, 6, 7, 8\} \cap \{1, 2, 7\} = \{1\}$$

and the characteristic set for case 1 from Table 2 is

$$K_A(1) = [(Wind, low)] \cap U \cap [(Temperature, high)]$$
$$= \{1, 2, 4, 5, 6\} \cap \{1, 2, 3, 4, 5, 6, 7, 8\} \cap \{1, 2, 3, 7, 8\} = \{1, 2\}.$$

All characteristic sets for Tables 1 and 2 are presented in Table 4. For a complete data set the characteristic set $K_B(x)$, where $x \in U$, is an equivalence class of the indiscernibility relation [12, 13].

## 3 Probabilistic Approximations

In our work we define probabilistic approximations based on the conditional probability of $X$ given $K_B(x)$, $Pr(X \mid K_B(x)) = \frac{|X \cap K_B(x)|}{|K_B(x)|}$ with $|Y|$ denoting the cardinality of set $Y$. Let $B$ be a subset of the attribute set $A$ and $X$ be a subset of $U$.

We further define three kinds of global probablistic approximations: singleton, subset and concept. A B-singleton probabilistic approximation of $X$ with the threshold $\alpha$, $0 < \alpha \leq 1$, denoted by $appr_{\alpha,B}^{singleton}(X)$, is defined as follows

$$\{x \mid x \in U,\ Pr(X \mid K_B(x)) \geq \alpha\}.$$

A B-subset probabilistic approximation of the set $X$ with the threshold $\alpha$, $0 < \alpha \leq 1$, denoted by $appr_{\alpha,B}^{subset}(X)$, is defined as follows

$$\cup\{K_B(x) \mid x \in U,\ Pr(X \mid K_B(x)) \geq \alpha\}.$$

A B-concept probabilistic approximation of the set $X$ with the threshold $\alpha$, $0 < \alpha \leq 1$, denoted by $appr_{\alpha,B}^{concept}(X)$, is defined as follows

$$\cup\{K_B(x) \mid x \in X,\ Pr(X \mid K_B(x)) \geq \alpha\}.$$

Global probabilistic approximations for the concept [(Trip, no)] from Table 1 are presented in Table 5.

**Table 5.** Global approximations for [(Trip, no)], Table 1

| Probabilistic approximations | | | |
|---|---|---|---|
| $\alpha$ | singleton | subset | concept |
| 1/3 | {2, 3, 5, 6, 7, 8} | {1, 2, 3, 5, 6, 7, 8} | {3, 5, 6, 7, 8} |
| 2/3 | {3, 5, 6, 7, 8} | {3, 5, 6, 7, 8} | {3, 5, 6, 7, 8} |
| 1 | {5, 6, 7} | {5, 6, 7} | {5, 6, 7} |

## 4 Experiments

In our experiments we used eight real-life data sets taken from the University of California at Irvine *Machine Learning Repository*. These data sets were modified by replacing 35% of existing attribute values by symbols of lost values, i.e., question marks. All data sets with lost values were edited, symbols of lost values were replaced by symbols of "do not care" conditions, i.e., by stars. Thus, for each data set, two data sets were created for experiments, one with missing attribute values interpreted as lost values and the other one as "do not care" conditions.

In our experiments we used the MLEM2 (Modified Learning from Examples Module, version 2) rule induction algorithm of the LERS (Learning from Examples using Rough Sets) data mining system [1, 3, 4].

Probabilistic rules were induced from modified data sets. For each concept $X$ and the set $Y$ equal to a probabilistic approximation of $X$ of a given type (singleton, subset or concept) a modified data set was created, see [9–11]. In this data set all cases from $Y$ had the same decision values as in the original data set, all remaining cases were labeled with a special, additional value. The LERS system, using the MLEM2 algorithm, was used to induce a rule set. Blocks of attribute-value pairs in the MLEM2 algorithm were modified, taking into account

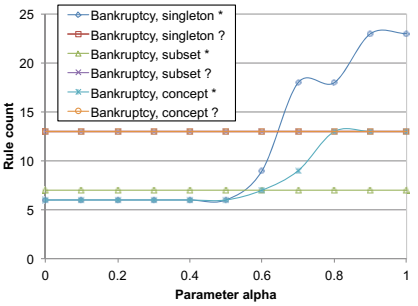



**Fig. 1.** Rule set size for the *bankruptcy* data set

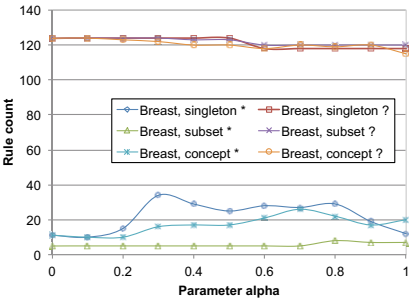



**Fig. 2.** Rule set size for the *breast cancer* data set

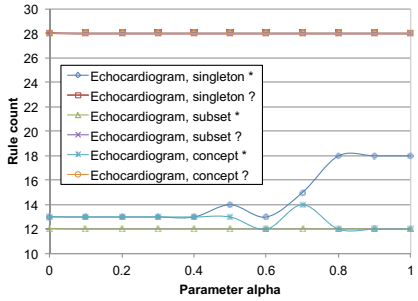



**Fig. 3.** Rule set size for the *echocardiogram* data set

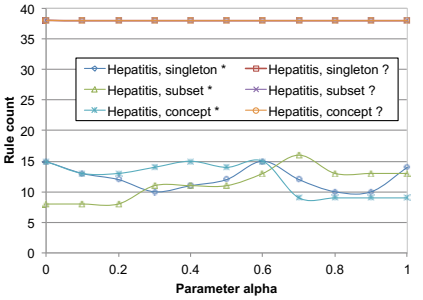



**Fig. 4.** Rule set size for the *hepatitis* data set

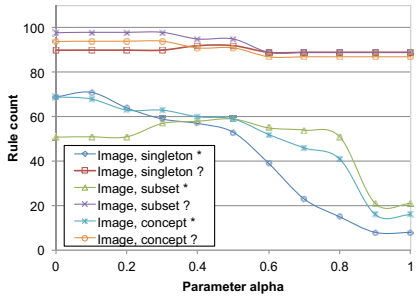



**Fig. 5.** Rule set size for the *image segmentation* data set

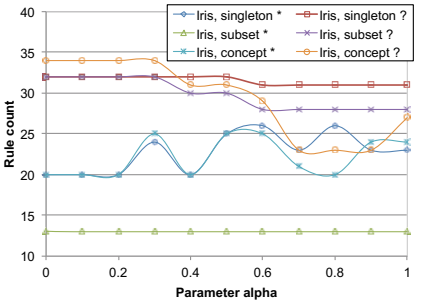



**Fig. 6.** Rule set size for the *iris* data set

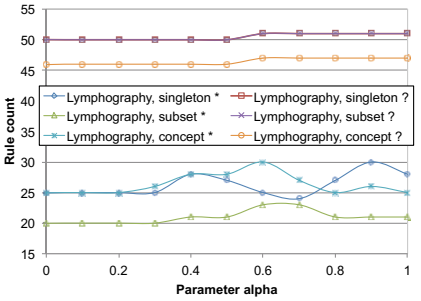



**Fig. 7.** Rule set size for the *lymphography* data set

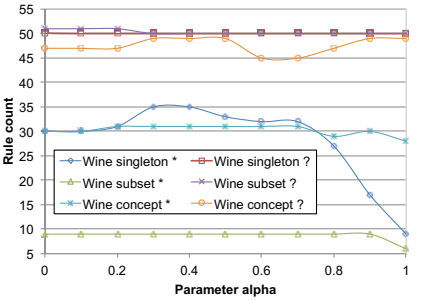



**Fig. 8.** Rule set size for the *wine recognition* data set

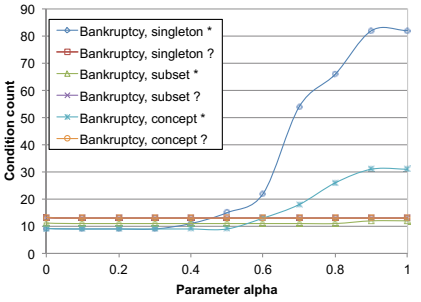



**Fig. 9.** Condition counts for the *bankruptcy* data set

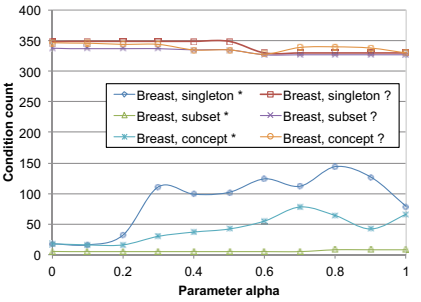



**Fig. 10.** Condition counts for the *breast cancer* data set

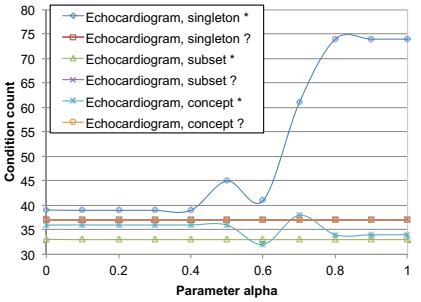



**Fig. 11.** Condition counts for the *echocardiogram* data set

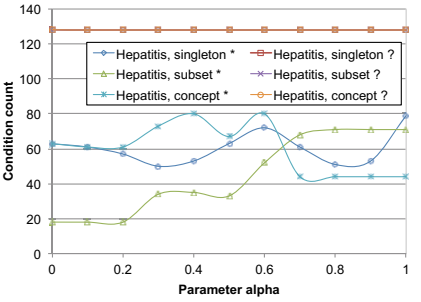



**Fig. 12.** Condition counts for the *hepatitis* data set

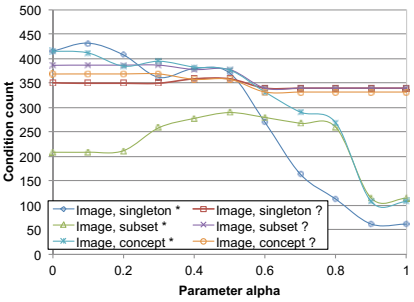



**Fig. 13.** Condition counts for the *image* data set

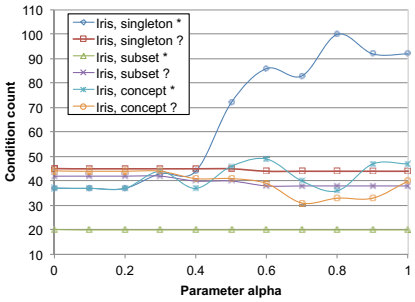



**Fig. 14.** Condition counts for the *iris* data set

missing attribute values. For the modified data set, only rules describing the concept $X$ survived, remaining rules were deleted. The aggregate rule set was combined from rule sets induced from all modified data sets.

For any data set we tested six methods of handling missing attribute values:

- singleton probabilistic approximation combined with lost values, denoted as Singleton, ?,
- singleton probabilistic approximation combined with "do not care" conditions, denoted as Singleton, $*$,
- subset probabilistic approximation combined with lost values, denoted as Subset, ?,
- subset probabilistic approximation combined with "do not care" conditions, denoted as Subset, $*$,
- concept probabilistic approximation combined with lost values, denoted as Concept, ?, and
- concept probabilistic approximation combined with "do not care" conditions, denoted as Concept, $*$.

As follows from [2], all six methods do not differ significantly (Friedman's test (5% significance level) in terms of the error rate.

Our main objective was to compare all six methods in terms of the complexity of rule sets. It is clear that for our data sets the method (Subset, $*$) provides smaller size of rule sets than all three methods associated with lost values: (Singleton, ?), (Subset, ?) and (Concept, ?). Additionally, the same method produces rule sets with smaller total number of conditions than all three methods associated with lost values.

Results of our experiments on the size of rule sets are presented in Figures 1–8. Six selected results on the total number of conditions (because of the space limit) are presented in Figures 9–14.

The method (Subset, $*$) provides smaller size of rule sets than (Singleton, $*$) and (Concept, $*$) for five out of eight data sets: Breast cancer, Echocardiogram,

Iris, Lymphography and Wine recognition and smaller total number of conditions for the same data sets (Wilcoxon test, 5% significance level was used for *Echocardiogram*).

Note that on some occasions the difference in performance is quite spectacular, for example, for the Breast cancer data set, (Subset $*$) method provides 5–7 rules (with $\alpha$ between 0.001 and 1) and with 5–8 conditions, while (Singleton, ?), (Subset, ?) and (Concept, ?) methods provide rule sets with 118–124 rules and 330–349 conditions. The error rate for (Subset, $*$) is between 28.52% and 29.90%, for all three methods associated with lost values, the error rate is between 27.44% and 29.90%.

Rule sets induced from data sets with "do not care" conditions are simpler, in general, than rule sets induced from data sets with lost values since for any data set, an attribute-value block for the data set with "do not care" conditions is a superset of the corresponding block (the same attribute-value pair) for the data set with lost values. The MLEM2 rule induction algorithm induces rules using these attribute-value blocks, so a rule induced from the data set with "do not care" conditions covers more cases than a rule induced from the data set with lost values.

## 5 Conclusions

For a given data set, all six methods of handling missing attribute values (using three kinds of global probabilistic approximations and two interpretations of missing attribute values) do not differ significantly with respect to the error rate [2]. However, as follows from our research presented in this paper, these methods differ significantly with respect to the complexity of rule sets; the simplest rule sets are induced using subset probabilistic approximations and missing attribute values interpreted as "do not care" conditions. Therefore, if we have a choice how to interpret missing attribute values, the best rule set would be induced by subset probabilistic approximations with missing attribute values interpreted as "do not care" conditions.

The focus of this work was a study of rule set complexity using different missing attribute interpretations and approximation methods while applying the same rule induction algorithm, MLEM2. Further investigation with other rule induction algorithms would be need in order to determine if the results are algorithm dependent.

## References

1. Clark, P.G., Grzymala-Busse, J.W.: Experiments on probabilistic approximations. In: Proceedings of the 2011 IEEE International Conference on Granular Computing, pp. 144–149 (2011)
2. Clark, P.G., Grzymala-Busse, J.W., Rzasa, W.: Mining incomplete data with singleton, subset and concept approximations (2013) (submitted for publication)
3. Grzymala-Busse, J.W.: A new version of the rule induction system LERS. Fundamenta Informaticae 31, 27–39 (1997)

4. Grzymala-Busse, J.W.: MLEM2: A new algorithm for rule induction from imperfect data. In: Proceedings of the 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 243–250 (2002)
5. Grzymala-Busse, J.W.: Rough set strategies to data with missing attribute values. In: Workshop Notes, Foundations and New Directions of Data Mining, in Conjunction with the 3rd International Conference on Data Mining, pp. 56–63 (2003)
6. Grzymala-Busse, J.W.: Characteristic relations for incomplete data: A generalization of the indiscernibility relation. In: Tsumoto, S., Słowiński, R., Komorowski, J., Grzymała-Busse, J.W. (eds.) RSCTC 2004. LNCS (LNAI), vol. 3066, pp. 244–253. Springer, Heidelberg (2004)
7. Grzymala-Busse, J.W.: Data with missing attribute values: Generalization of indiscernibility relation and rule induction. Transactions on Rough Sets 1, 78–95 (2004)
8. Grzymala-Busse, J.W.: Generalized parameterized approximations. In: Yao, J., Ramanna, S., Wang, G., Suraj, Z. (eds.) RSKT 2011. LNCS, vol. 6954, pp. 136–145. Springer, Heidelberg (2011)
9. Grzymala-Busse, J.W.: Generalized probabilistic approximations. Transactions on Rough Sets 16, 1–16 (2013)
10. Grzymala-Busse, J.W., Marepally, S.R., Yao, Y.: A comparison of positive, boundary, and possible rules using the MLEM2 rule induction algorithm. In: Proceedings of the 10th International Conference on Hybrid Intelligent Systems, pp. 7–12 (2010)
11. Grzymala-Busse, J.W., Yao, Y.: Probabilistic rule induction with the LERS data mining system. International Journal of Intelligent Systems 26, 518–539 (2011)
12. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
13. Pawlak, Z., Grzymala-Busse, J.W., Slowinski, R., Ziarko, W.: Rough sets. Communications of the ACM 38, 89–95 (1995)
14. Pawlak, Z., Skowron, A.: Rough sets: Some extensions. Information Sciences 177, 28–40 (2007)
15. Pawlak, Z., Wong, S.K.M., Ziarko, W.: Rough sets: probabilistic versus deterministic approach. International Journal of Man-Machine Studies 29, 81–95 (1988)
16. Ślęzak, D., Ziarko, W.: The investigation of the bayesian rough set model. International Journal of Approximate Reasoning 40, 81–91 (2005)
17. Yao, Y.Y.: Probabilistic rough set approximations. International Journal of Approximate Reasoning 49, 255–271 (2008)
18. Yao, Y.Y., Wong, S.K.M.: A decision theoretic framework for approximate concepts. International Journal of Man-Machine Studies 37, 793–809 (1992)
19. Ziarko, W.: Probabilistic approach to rough sets. International Journal of Approximate Reasoning 49, 272–284 (2008)

# A Possibilistic View of Binomial Parameter Estimation

Gilles Mauris

LISTIC, Univ. Savoie, F-74000 Annecy
BP 80 439, 74 944 Annecy Le Vieux, France
gilles.mauris@univ-savoie.fr

**Abstract.** This paper deals with the possibility roots of binomial parameter interval estimation. It shows that conventional probability methods consist to obtain confidence intervals representing *de dicto* parameter uncertainty from coverage intervals representing *de re* uncertainty of observed samples. We relate the different types of coverage intervals to equivalent *de re* possibility distributions whose lead after inversion to *de dicto* possibility distributions corresponding to the stacking up of all confidence intervals at all levels. The different choices for the centre of the intervals corresponds to the different existing methods, in the same vein a novel one centred on the mean is proposed.

**Keywords:** Possibility theory, binomial parameter estimation, coverage intervals, confidence intervals.

## 1 Introduction

The first distribution introduced in probability is the Bernoulli ones that is a simple discrete distribution taking only two outcomes: *0* (failure) and *1* (success). The probability that the considered variable takes the value *1* is conventionally denoted $p$. If a Bernoulli trial is repeated independently $n$ times, the random variable $S_n$ defined by the sum of successes follows a binomial distribution denoted $Bin(n,p)$. The binomial distribution is frequently used to model a lot of issues involving decision from samples: controlling failures in components' production, medical and biological tests, demographic analysis, opinion poll, games, ... But, beyond these practical interests, the binomial proportion estimation is fundamental in the very definition of probability. Indeed, the estimation of $p$ is at the root of the justification of the unknown probability estimation by the observed realised frequency on a large sample, thanks to the weak law of large numbers established by Jacob Bernoulli [1]. Therefore, it is important to study the practical and theoretical sides of what the possibility theory [2][3] has to say on the binomial parameter estimation. This is the purpose of this article that considers the existing probability approaches under a possibilistic angle from probability-possibility transformations [4][5][6] identifying the coverage intervals of a probability distribution to alpha-cuts of the equivalent possibility distributions.

First, fundamental notions of probability estimation and possibility theories are recalled. We point out the fact that the estimation of the parameter $p$, which is fixed but unknown, is characterized by a so called *de dicto* uncertainty (because it concerns

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 396–405, 2014.

knowledge uncertainty [7]) that is deduced in one way or another from a so-called *de re* uncertainty (because it concerns observed things) issued from the sample. Secondly, we present a possibility-theory-based view of the conventional probability approaches used to build confidence intervals for a binomial parameter [8][9]. A new estimation method (to our knowledge) based on asymmetrical coverage intervals centred on the mean is then proposed. Finally, we conclude on the unified framework proposed by the possibilistic view and on the new estimation methods that are suggested by the possibility view.

## 2 Fundamental Probability Estimation Notions

In this section, we present first in an intuitive way the fundamental binomial estimation concepts already involved in the pioneers' works. Then, we propose modern formulation of these concepts and relate them to the notion of coverage and confidence intervals.

### 2.1 *de re* and *de dicto* Uncertainties

The estimation of the binomial parameter $p$ is at the root of the justification of the estimation of an unknown probability by the observed frequency $F_n = \frac{S_n}{n}$ on a large sample of size $n$, thanks to the weak law of large numbers established by Jacob Bernoulli since the beginning of probability [1]:

$$\forall \varepsilon > 0 \qquad P(|F_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \tag{1}$$

This theorem expresses the probability concentration increase phenomenon as $n$ increases. In fact this theorem allows one to deduce information from dispersed observed data. Indeed, as $n$ increases, the fluctuations of $F_n$ around $p$ are decreasing and we become quasi-sure that $F_n = p$ .

Reading the equation (1) in a direct way allows from the knowledge of $p$ to make a prevision (with a considered risk) of the fluctuation of $F_n$ , *i.e.*, to define an interval in which $F_n$ is to be found within at least a definite probability (thus called a coverage interval). In a reciprocal way, the equation (1) allows to determine an interval depending upon $F_n$ containing the unknown fixed parameter $p$ with a probability level (thus called confidence interval). In fact, the more concentred is the probability the narrower are the coverage intervals, and the narrower are the confidence intervals obtained by inversion. At the limit for $n$ infinite, $F_n$ fluctuates no more (it becomes a Dirac distribution), and we can deduce with complete certainty that $F_n = p$ .

The preceding considerations highlight a fundamental point: the obtained knowledge on the fixed unknown $p$ is characterized by a *de dicto* uncertainty that is related to the *de re* uncertainty issued from the fluctuations of the observed sample.

### 2.2 Formal Definition of Coverage Intervals

Let us first define a discrete random variable $X$ on a probability space $(\Omega, \mathcal{T}, P)$ composed of a sample space $\Omega$; a set of subset of events $\mathcal{T}$; a function $P$ returning an event's probability as $X : \Omega \to \mathbb{R}, \omega \mapsto X(\omega)$ such that: $X(\Omega) = \{X(\omega), \omega \in \Omega\}$ is a countable subset of $\mathbb{R}$.

$\forall x_k \in X(\Omega), A_k = \{\omega \in \Omega, X(\omega) = x_k\}$ belongs to the family $\mathcal{T}$ of events having a probability assigned by $P$; $X(\omega) = x_k$ is called a realisation of the random variable. The application $X$ allows to transport the probability $P$ from $\Omega$ onto $\mathbb{R}$; The $P(X = x_k)$'s are considered as punctual masses and the probability of a subset of $\mathbb{R}$ is defined as the sum of its punctual masses. The set function $P_X$ defined by $p_k = P_X(\{x_k\}) = P(A_k) = P(X = x_k)$ is called the distribution function of the random variable $X$.

As seen previously, the realisations of a random variable fluctuate and thus are dispersed, therefore it is useful to know the intervals in which the realisations are included with a specified proportion. This is formalized by the notion of coverage interval of level $1-\alpha$ $(\alpha \in [0,1])$ of a random variable $X$ defined as an interval of the form $\left[G_X^{-1}(\beta), G_X^{-1}(\beta + 1 - \alpha)\right]$ where $G_X^{-1}$ is the inverse cumulative distribution function of $X$ (also called quantile function). These coverage intervals are not unique, different choices for $\beta \in [0,1]$ lead to different types of coverage intervals: lower unilateral ($\beta = 0$), upper unilateral ($\beta = \alpha$), bilateral ($\beta = \frac{\alpha}{2}$). Generally, they are built about central points such as the mean, the mode, the median. The coverage intervals are sometimes called prediction intervals because as a coverage interval of level $1-\alpha$ contains $(1-\alpha)\%$ of the population, there is $(1-\alpha)\%$ chance that an observation coming from this population falls in this interval. Note that the coverage intervals for a discrete random variable do not always exist for all the level due to the discreteness of the realisations.

Finally, let us point out the deterministic nature of the coverage intervals which have definite bounds, and thus the *de re* nature of the uncertainty they model. Indeed, they reflect the observation fluctuations (due to the very phenomenon or to the way they are observed). Therefore, to increase the number of observations does not decrease the *de re* uncertainty but simply improves the validity of its representation according to the observed random phenomenon.

### 2.3 Formal Definition of Theoretical and Realised Confidence Intervals

The notion of confidence interval is more subtle because it aims at defining the uncertainty of a fixed unknown parameter (i.e. without fluctuations), and thus corresponds to a *de dicto* uncertainty characterizing the knowledge declared. The main difficulty for determining confidence intervals comes from the fact that as $p$ is unknown, hence there are multiple probability measure $P$ we could consider for the

probability space $\left(\Omega_n = \{0,1\}^n, \mathcal{F}_n = 2^{\Omega_n}\right)$. In fact, only a family $\left(P_p\right)_{p\in[0,1]}$ of probability measures defined by $P_p(\{\omega\}) = p^{S_n(\omega)}(1-p)^{n-S_n(\omega)}$ can be soundly considered with $S_n(\omega) = \sum_{i=1}^{n} X_i(\omega)$ *and* $X_i(\omega) = 0 or 1$. We have thus a so called statistical model $\left(\Omega_n, \mathcal{F}_n, \left(P_p\right)_{p\in[0,1]}\right)$ with for every $p \in [0,1]$ the distribution of $S_n$ under $P_p$ that follows a binomial distribution *Bin(n,p)*. The theoretical confidence intervals are then defined in the following way. Let $\left(\Omega_n, \mathcal{F}_n, \left(P_p\right)_{p\in[0,1]}\right)$ be the statistical model and $X = (X_1, X_2, ..., X_n)$ a sample coming from this model, a theoretical confidence interval for $p$ of confidence level at least $1-\alpha$ is a closed interval with random variable bounds *L(X)* and *U(X)* satisfying [8]:

$$\inf_{p\in[0,1]} P_p(L(X) \le p \le U(X)) \ge 1-\alpha$$

It is very important to note that the function $\inf_{p\in[0,1]} P_p$ defined on $\mathcal{F}_n$ by $A \mapsto \inf_{p\in[0,1]} P_p(A)$ does not provide a probability but a lower bound of probability. That is why we speak about a level of confidence and not about a level of probability (we will see in the next sub-section that this level can be related to possibility/necessity notions). Further, when we replace the random variable *X* with its realisation we obtain for *p* realised (or numerical) confidence intervals ($\left[L(f_n), U(f_n)\right]$ ($f_n$ is the observed frequency) with deterministic bounds. The event $p \in \left[L(f_n), U(f_n)\right]$ is either true or false, but it is not subject to fluctuations, and thus cannot be interpreted as a frequentist probability. In fact, the theoretical confidence interval is a procedure (often called estimator) which when applied an infinite times leads to a success rate at least equal to the level of confidence. For example, it is sound to say before making the observations that for *100* confidence intervals of level *90%* realised in similar conditions, *90* of them will contain the unknown fixed parameter. But our purpose is to have information about the uncertainty of *p* after having made the observations, *i.e.*, once having obtained a realised confidence interval $\left[L(f_n), U(f_n)\right]$. In fact, this realised confidence interval conveys a *de dicto* uncertainty. Indeed, it seems logical to us to transfer the uncertainty (i.e. the level of confidence) of the theoretical confidence interval to the realised confidence interval. Indeed, in itself (in absence of other information) the realisation does not change in our opinion the degree of probability (subjective here assigned to the event $p \in \left[L(f_n), U(f_n)\right]$. This transfer is coherent with the notion of bet at the basis of subjective probability. Indeed, if we are ready to bet *90* euros against *10* euros that a random interval of level *90%* contains the unknown parameter *p*, we are also ready to bet *90* euros that the unknown parameter is within the realised confidence interval of level *90%*. We meet again here the idea of prediction conveyed by a realised confidence interval. Finally, let us point about the random nature of theoretical

confidence intervals and the *de dicto* nature of the uncertainty (obtained here by transfer) conveyed by realised confidence intervals. Indeed, they are a reflection of the knowledge incompleteness about $p$ that comes from the limited number of observations and of their fluctuations. Thus, the uncertainty on the knowledge of $p$ could be reduced by increasing the number of observations.

## 3 Possibility Distribution versus Coverage and Confidence Intervals

As mentioned above, beyond vocabulary and definition difficulties, the probability interpretation of confidence intervals is not so clear. We hope that the possibility theory based view of coverage and confidence intervals exposed in this section will allow to better understand and to reconcile historical and current points of view.

### 3.1 Basics of Possibility Theory

A fundamental notion of the possibility theory [3] is the possibility distribution, denoted $\pi$. Here, we consider possibility distributions defined on the real line, i.e. $\pi$ is an upper semi-continuous mapping from $R$ to the unit interval such that $\pi(x) = 1$ for some $x$ belonging to $R$. A possibility distribution generates two non-additive set functions [3]: possibility measure $\Pi$ and a necessity measure $N$ as follows. $\forall A \subset R, \Pi(A) = \sup_{x \in A} \pi(x)$ and $\forall A \subset R, N(A) = 1 - \Pi(\overline{A}) = \inf_{x \notin A} (1 - \pi(x))$

The possibility measure $\Pi$ satisfies: $\forall A, B \subset R, \Pi(A \cup B) = \max(\Pi(A), \Pi(B))$

The necessity measure $N$ satisfies: $\forall A, B \subset R, N(A \cap B) = \min(N(A), N(B))$

To qualify the informativeness of a possibility distribution, the concept of specificity is useful. A possibility distribution $\pi_1$ is called more specific (i.e. thinner in a broad sense) than $\pi_2$ as soon as $\forall x \in R, \pi_1(x) \leq \pi_2(x)$ (fuzzy set inclusion). The more specific $\pi$, the more informative it is. If $\pi(x) = 1$ for some $x$ and $\pi(y) = 0$ for all $y \neq x$, then $\pi$ is totally specific (fully precise and certain knowledge), if $\pi(x) = 1$ for all $x$ then $\pi$ is totally non specific (complete ignorance). In fact, a numerical degree of possibility can be viewed as an upper bound to a probability degree [10]. Namely, with every possibility distribution $\pi$ one can associate a non-empty family of probability measures dominated by the possibility measure: $\mathcal{P}(\pi) = \{P, \forall A \subset R, P(A) \leq \Pi(A)\}$. This provides a bridge between probability and possibility.

### 3.2 Relationships with Coverage Intervals

For the same probability density function and for the same probability level, we can have different types of coverage intervals. Indeed, we can impose one of the bounds or the centre or another point of the coverage intervals. A coverage interval is said

optimum if its length is the smallest among all the coverage intervals of the same level. For all types, coverage intervals are nested versus the probability level, and thus can be represented as a whole by a possibility distribution. Indeed, an unimodal numerical possibility distribution can be viewed as a nested set of intervals, which are the $\alpha$-cuts of $\pi:[\underline{x}_\alpha,\overline{x}_\alpha]=\{x,\pi(x)\geq\alpha\}$ and these α-cuts can be identified with the coverage intervals of probability level $\beta=1-\alpha$ of a random variable *X*, i.e. an interval that contains a portion of *X* with probability $\geq\beta$. This approach has been used in previous works [5][6] to define a probability-possibility transformation.

One important result is that for any random variable *X* having a cumulative distribution function *G*, and for any central point *c*, a possibility distribution of type *1* defined by [6] :

$\pi_x^{1c}(x)=G(x)+1-G(h(x))=\pi_x^{1c}(h(x))$ (with $h:[-\infty,c]\to[c,+\infty]$ a decreasing function such that $h(c)=c$) is an unimodal distribution with mode *c* satisfying : $\forall A,\prod_X^{1c}(A)\geq P_X(A)$. If *X* has an unimodal density *f* (*i.e. f* is strictly increasing before the mode *M* and strictly decreasing after *M* ), an optimal equivalent possibility distribution is obtained by : $\forall x\in[-\infty,M],h(x)=\{\inf\ y\geq M\,|\,f(x)=f(y)\}$.

Another new way we propose here is to represent coverage intervals by a possibility distribution of type *2* defined as follows:

$$\pi_x^{2c}(x)=\frac{G(l(x))}{G(c)}\quad for\quad x\leq c\quad and\quad \frac{1-G(r(x))}{1-G(r(c))}\quad for\quad x\geq c$$

with *r* and *l* being respectively increasing and decreasing functions defined on $[-\infty,c]et[c,+\infty]$ and such that $l(c)=r(c)=c$. The possibility distribution $\pi_x^{2c}$ is unimodal with mode *c* and it satisfies: $\forall A,\prod_X^{2c}(A)\geq P_X(A)$.

*Proof*: $\pi_x^{2c}(c)=1$ since $l(c)=r(c)=c$, then if $A\supset c,\prod_X^{2c}(A)=1\geq P_X(A)$, and if $c\notin A$ and $\sup(x\in A)<c$ ,then $\pi_x^{2c}(x)=\frac{G(l(x))}{G(c)}\geq G(l(x))$ because $G(c)<1$ ,therefore $\prod_X^{2c}(A)\geq P_X(A)$. If $c\notin A$ and $\inf(x\in A)>c$, $\pi_x^{2c}(x)=\frac{1-G(r(x))}{1-G(r(c))}\geq 1-G(r(x))$ □.

Finally, let us remark that if the probability distribution is symmetric about *c,* then if *l(x))=r(x)=x*, we have $\pi_x^{1c}(x)=\pi_x^{2c}(x)=\min(2G(x),2(1-G(x))$.

### 3.3 Relationships with Realised Confidence Intervals

As explained before, the realised confidence intervals are such that $\inf_{p\in[0,1]}P_p(L(x)\leq p\leq U(x))\geq 1-\alpha$, i.e. $\inf_{p\in[0,1]}P_p([L(x),U(x)])\geq 1-\alpha$, therefore the confidence level of the realised confidence interval is a necessity degree. If we dispose of a family of realised nested confidence intervals $\{I_1,I_2,...,I_m\}$

$I_i \subset I_{i+1}, i = 1,...,m$, every $I_i$ having a confidence level of $\lambda_i$ defined as a necessity degree then the equivalent possibility distribution is defined by [10]:

$\pi_{X=x}(x) = \min_{i=1,...,m} \max(1-\lambda_i, I_i(x))$ where $I_i(x)$ is the characteristic function of $I_i$.

The preceding equation expresses in fact the conjunction of the possibility distributions equivalent to each realised confidence interval of each level. The resulting possibility distribution corresponds to the more specific *de dicto* possibility distribution according to the available information.

Another way to obtain directly (without considering the realisation of random variables) *de dicto* confidence intervals is to consider Fisher inference, leading to fiducial confidence intervals [11]. According to us, Fisher idea may be expressed in a possibility setting by considering *possible* at a definite level all values of the parameter *p* whose corresponding coverage intervals at the same level contains the observation. The fiducial confidence intervals are thus clearly related by inversion to coverage intervals. Interestingly, in a lot of situations (particularly when the distributions are increasing with the parameter as it is the case for the binomial distribution) the realised confidence intervals are equal to the fiducial confidence intervals, and therefore we will not discuss further here the controversial aspects of frequentist and fiducial inferences.

# 4 Methods for Building Binomial Confidence Intervals

In this section we recall first the main properties of the binomial distribution, and then we present the usual methods for building binomial confidence intervals [9] in a possibility setting. Finally a new derived method is exposed.

## 4.1 Binomial Distribution Properties

Let us denote $S_n$ the random variable associated to the repetition of *n* identical and independent Bernoulli trials having *p* as success probability. $S_n$ follows a binomial distribution *Bin(n,p)* defined by: $P(S_n = k) = C_n^k p^k (1-p)^{n-k}$. This binomial distribution is asymmetric (except for *p=0.5*) and its mean is $E = np$. If $(n+1)p$ is not an integer, the distribution strictly increases on $[0, \lfloor (n+1)p \rfloor]$ and strictly decreases on $[\lfloor (n+1)p \rfloor, n]$, and thus it has only one mode $M = \lfloor (n+1)p \rfloor$, and the mean is equal to the mode $E = M = np$. If $(n+1)p$ is an integer the distribution strictly increases on $[0,(n+1)p-1]$ and strictly decreases on $[(n+1)p, n]$; thus it has two modes $M = (n+1)p-1$ *and* $(n+1)p$ and the mean $E = np \in [M-1, M]$ is not reachable. In all cases, the distance between the mean and the mode is such that $|M-E| < 1$. The median is defined by $m = \lfloor np \rfloor ou \lceil np \rceil \pm 1$. Note that the reason of non monotony observed sometimes when inverting coverage intervals is due to fact that the expressions of the mode and of the median are varying with *n* and *p*.

### 4.2 Wald-DeMoivre-Laplace Method

It is the gold standard method (popularized by Wald) based on the approximation of the binomial distribution by a normal distribution (established by Laplace on the basis of DeMoivre derivations). The estimator $\hat{p}$ (obtained from the observations) of the proportion $p$ follow approximately a normal distribution $N(\hat{p},\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$, and thus $\hat{p}-p$ follows approximately a normal distribution with a mean equal to 0 and a standard deviation equal to $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ (here we have replaced the unknown standard deviation with its sample estimation; the Wilson score method [9] consider the true standard deviation). The distribution of random variable $\hat{p}-p$ is what is called an asymptotic pivotal distribution though it does not depend upon the unknown parameter $p$. From inverting the *de re* possibility distribution of type *1* around *0* equivalent to this normal probability distribution, confidence intervals can be easily deduced. Let us note that the normal approximation is more or less good and that it does not grant the level of confidence (see figure 1). The fact that no exact pivotal distribution (i.e. non-asymptotic) exists for a discrete distribution (unlike continuous distributions) makes the building of confidence intervals a difficult task.

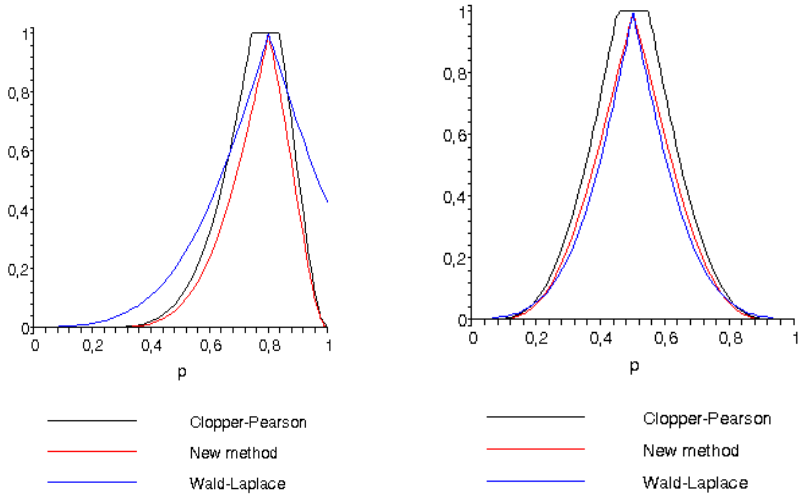



**Fig. 1.** *de dicto* possibility distributions for *8* (left) and *5* (right) successes for *10* trials

### 4.3 Clopper-Pearson Method

The Clopper-Pearson [9] approach consists in inverting the *de re* possibility distribution of type *2* about the median equivalent to the binomial distribution ($G_p$ denotes the cumulative binomial distribution with a success proportion $p$).

The expression of the *de dicto* possibility distribution corresponding to an observation of $k$ failures is thus:

$$\pi^{CP}(p,k) = \min(2G_p(k/n), 2(1-G_p(k/n), 1) = \pi^{2med}_{G_p}(k/n)$$

An important point is that this *de dicto* possibility distribution is unimodal but it is not the most specific. Indeed, Blaker [9] by inverting the *de re* possibility distribution of type *1* about the median obtained a possibility distribution included in the former. But in a few situations the Blaker's possibility distribution of $p$ is not unimodal

### 4.4 Wilson-Sterne, Crow and Clunies-Ross Methods

In order to obtain the more specific *de dicto* possibility distribution, Wilson (1942) considered the more specific de re possibility distribution, i.e. the one of type 1 about the mode. This method is more known under the name of Sterne method (1954) or of likelihood minimum [9]. The possibility distribution has the following expression:

$$\pi^{St}(p,k) = G_p(k'/n) + (1-G_p(k/n)), p \le k/n \quad \text{with} \quad k = h(k')$$
$$\pi^{St}(p,k) = G_p(k/n) + (1-G_p(k''/n)), p \ge k/n \quad k'' = h(k)$$

Thus: $\pi^{St}(p,k) = \pi^{1mod}_{G_p}(k/n)$

This method does not ensure to obtain an unimodal distribution. A modification proposed by Crow (1956) prevents of having disjoint coverage intervals, but it leads sometimes after inversion to non nested confidence intervals. A more sound approach is the one of Clunies–Ross (1958)[9] that proposed to fill in the holes between the disjoint intervals, leading thus to an unimodal possibility distribution; the price to pay is a less specific possibility distribution.

### 4.5 A New Method

An approach we have not seen in the literature is to invert the *de re* possibility distribution of type 2 about the mean leading to :

$$\pi^{new}(p,k) = \min\left(\frac{G_p(k/n)}{G_p(p)}, \frac{1-G_p(k/n)}{1-G_p(p)}, 1\right) = \pi^{2mean}_{G_p}(k/n)$$

This *de dicto* possibility distribution is unimodal and has the advantage to be centred about the maximum likelihood estimates.

The figure 1 at the end of the paper illustrates cases where respectively *8* and *5* successes have been observed for *10* trials.

## 5 Conclusion

In this paper, we have highlighted the relationships between conventional parameter interval estimation methods and possibility distributions. Indeed, the most used probability methods amount to invert the *de re* possibility distribution equivalent to

all coverage intervals built about different centres (mode, median, mean) to obtain a *de dicto* possibility distribution equivalent to the fiducial confidence intervals of all confidence levels. A new method based on a new probability-possibility transformation centred about the sample mean of the binomial proportion (i.e. the maximum likelihood estimation) has been exhibited. In perspectives, the Bayesian approach of binomial parameter estimation could be cast in the possibility framework by transforming the posterior probability distribution into a possibility distribution.

## References

1. Hald, A.: A History of Parametric Statistical Inference from Bernoulli to Fisher, pp. 1713–1935. Springer (2008)
2. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1(1), 3–28 (1978)
3. Dubois, D., Prade, H.: Possibility Theory. Plenum, New York (1988)
4. Dubois, D., Prade, H., Sandri, S.: On possibility/probability transformations. Fuzzy Logic, 103–112 (1993)
5. Mauris, G., Lasserre, V., Foulloy, L.: A fuzzy approach for the expression of uncertainty in measurement. Int. Journal of Measurement 29(3), 165–177 (2001)
6. Dubois, D., Foulloy, L., Mauris, G., Prade, H.: Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities. Reliable Computing 10(4), 273–297 (2004)
7. Boudot, M.: Probabilité et logique de l'argumentation selon Jacques Bernoulli. In Philosophie et Logique, Presses de l'Université Paris Sorbonne, 15–37 (2009)
8. Tassi P.: Méthodes Statistiques. Economica, 3rd edn., Paris (2004)
9. Brown, L., Tony Cai, T., DasGupta, A.: Interval estimation for a binomial proportion. Statistical Science 16(2), 101–133 (2011)
10. Dubois, D., Prade, H.: When upper probabilities are possibility measures. Fuzzy Sets and Systems 49, 65–74 (1992)
11. Wang, H.Y.: Fiducial intervals what are they? The American Statistician 54(2), 105–111 (2000)

# A New Approach to Economic Production Quantity Problems with Fuzzy Parameters and Inventory Constraint

József Mezei[1] and Kaj-Mikael Björk[1,2]

[1] IAMSR, Åbo Akademi University, Turku
[2] Arcada University of Applied Sciences, Helsinki
{jmezei,kaj-mikael.bjork}@abo.fi

**Abstract.** In this paper, we will develop a new multi-item economic production quantity model with limited storage space. This new model will then be extended to allow for fuzzy demand and solved numerically with a non-linear programming solver for two cases: in the first case the optimization problem will be defuzzified with the signed distance measure and in the second case, the storage constraint needs to be fulfilled, only to a certain degree of possibility. Both cases are solved and illustrated with an example.

**Keywords:** Economic Production Quantity, Triangular fuzzy numbers, Inventory constraint, Signed distance, Chance constrained optimization.

## 1 Introduction

Even with more than 100 years of EOQ (Economic Order Quantity) development, current stream of new findings and results do not tend to decrease. Even if the first model by Harris [8] was very simple, it has been very popular in industry and also an inspiration to many researchers. In this basic model the order size needed to be determined given holding costs, order setup costs and annual demand. This model has been altered in many ways to capture more complex and realistic situations in the industry. For instance, the EPQ (Economic Production Quantity) solved a problem where the product is produced to stock, also multi item, storage capacity limitation and so on is further extensions of the basic model.

These additions may be very crucial, even to the extent of only having storage capacity for one weeks production (this was the case in a Nordic plywood production facility that we have collaborated with). It is obvious that we will produce to stock in the process industry environment. In these settings we need the EOQ-models with some proper extensions. The uncertainties in the process industry can sometimes be measured probabilistically, but sometimes data is not enough and therefore fuzzy measures may be needed, c.f [3,5]. There have also been a lot of research contributions in this line of research. For instance [3] solved an EOQ model with backorders and infinite replenishment lead time

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 406–415, 2014.

with fuzzy lead times. However, sometimes the environment may be more stable, and only a few things may be uncertain. These fuzzy uncertainties may come from the fact that the demand may be uncertain, but still reliable data is not found to make justified probabilistic statements. This case is tackled in our paper for a special case of inventory constraints. Often is desirable to try to solve the EOQ-models with their extensions analytically through the solution of the derivatives (as also done originally by Harris, [8]). There are also other optimization approaches used in the EOQ literature. If the uncertainties in the EOQ-models can be modeled stochastically (as done in [9]), the track of probabilistic models should be conducted, but this is not always possible in the process industry. For the uncertainties relevant to this paper it is better to use fuzzy numbers instead of probabilistic approaches ([17,18]). In the line of research of fuzzy EOQ-models, there are contributions for instance like Chang [6], who worked out fuzzy modifications of the model of [13], which took the defective rate of the goods into account. Ouyang and Wu and [11] Ouyang and Yao [12] solved an EOQ-model with the lead times as decision variables as well as the order quantities. Taleizadeh and Nematollahi [15] presented again an EOQ-model with a final time horizon, with perishable items, backordering and delayed payments. Sanni and Chukwu [14] did a EOQ-model with deteriorating items, ramp-type demand as well as shortages. This paper has a track of research development behind. Already Björk and Carlsson [3] solved analytically an EOQ problem with backorders, with a signed distance defuzzification method. Björk [1] solved again a problem with a finite production rate and fuzzy cycle time, which was extended in [2] to a more general fuzzy case. The approach used in this paper is novel since there are no papers (to our knowledge) that focus on the realistic modeling of inventory constraints. Our solution methodology is to one part similar also to Björk and Carlsson [4] and Björk [1], , where the fuzzy model is defuzzified using the signed distance method [16], however, the solution is here not found through the derivatives, but numerically, since our fuzzy problem is more difficult to solve. This paper extends the results in the recent publication by Björk [2] with the limited storage capacity restriction with a more complex, but much more realistic inventory space constraint model. In addition, we consider not only the crisp case, but also the case of chance constrained formulation (in the fuzzy sense) of the storage limitations. The rest of the paper is structured as follows. First we will explain some preliminaries, then we will present the crisp case, after which we allow the demand to be fuzzy. Finally we solve the model both with defuzzification method as well as introducing fuzzy necessity constraints. Finally we will show this with an example as well as make some concluding remarks.

## 2 Preliminaries

In this section we introduce the necessary definitions and notations that are necessary for developing and solving our new model. We focus on fuzzy numbers and possibilistic chance constrained programming.

### 2.1 Fuzzy Numbers

Fuzzy sets have been introduced by Zadeh [17] to represent uncertainty different from randomness. In this paper, we employ fuzzy sets to model incomplete information inherent in many real world applications of inventory management. The most used special case of fuzzy sets is the family of triangular fuzzy numbers.

**Definition 1** *Consider the fuzzy set* $\tilde{A} = (a, b, c)$ *where* $a < b < c$ *and defined on* $\mathbb{R}$*, which is called a triangular fuzzy number, if the membership function of* $\tilde{A}$ *is given by*

$$\mu_{\tilde{A}}(x) = \begin{cases} \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ \frac{c-x}{x-b} & \text{if } b \leq x \leq c \\ 0 & \text{otherwise} \end{cases}$$

In order to find non-fuzzy values for the model, we need to use some distance measures, and we will use the signed distance [16]. Before the definition of this distance, we need to introduce the concept of $\alpha$-cut of a fuzzy set.

**Definition 2** *Let* $\tilde{B}$ *a fuzzy set on* $\mathbb{R}$ *and* $0 \leq \alpha \leq 1$*. The* $\alpha$*-cut of* $\tilde{B}$ *is the set of all the points* $x$ *such that* $\mu_{\tilde{B}}(x) \geq \alpha$*, i.e.* $\tilde{\tilde{B}}(\alpha) = \{x | \mu_{\tilde{B}}(x) \geq \alpha\}$ *.*

Let $\Omega$ be the family of all fuzzy sets $\tilde{B}$ defined on $\mathbb{R}$ for which the $\alpha$-cut $\tilde{B}(\alpha) = [\tilde{B}_l(\alpha), \tilde{B}_u(\alpha)]$ exists for every $0 \leq \alpha \leq 1$, and both $\tilde{B}_l(\alpha)$ and $\tilde{B}_u(\alpha)$ are continuous functions on $\alpha \in [0, 1]$.

**Definition 3** *For* $\tilde{B} \in \Omega$ *define the signed distance of* $\tilde{B}$ *to* $\tilde{0}$ *as*

$$d(\tilde{B}, \tilde{0}) = \frac{1}{2} \int_0^1 (\tilde{B}_l(\alpha) + \tilde{B}_u(\alpha)) \mathrm{d}\alpha$$

### 2.2 Chance Constrained Programming

Chance constrained programming, originally introduced in probabilitic environment by Charnes and Cooper [7], is a widely-used method to handle uncertain parameters in optimization problems. The original approach was later modified to incorporate fuzzy parameters and possibility and necessity measures [10]. According to this approach, it is not necessary to use any defuzzification method, the extent to which the constraints of the models are satisfied in terms of possibility or necessity are calculated.

Possibility measure [19] is a maxitive normalized monotone measure, i.e.

$$Pos\left(\bigcup_i B_i\right) = \sup_i Pos(B_i).$$

where $\{B_i\}$ is any family of sets in the universe of discourse. The dual measure of possibility, termed as necessity, is defined as:

$$Nec(B) = 1 - Pos(B^C).$$

We can consider fuzzy numbers as possibility distributions on the real line using the formula

$$Pos(C \subset \mathbb{R}) = \sup_{x \in C} \mu_B(x),$$

where $\mu_B(x)$ is the membership function of the fuzzy number $B$. In this paper we will calculate the possibility of the fulfilment of constraint with the left-hand side being a fuzzy expression and the right-hand side as a crisp number (size of available storage). As crisp numbers are special cases of fuzzy numbers we can use the following formula for $A \in \mathbb{R}$:

$$Pos(B \leq A) = \sup \{\mu_B(x) \mid x \leq A\}.$$

## 3 EPQ Model with Fuzzy Parameters and Storage Constraint

In this section, we are first going to present the crisp and fuzzy models, and two approaches for solving the fuzzy formulation. The parameters and variables (can be assumed strictly greater than zero) in the classical multi-item EPQ model with shared cycle time and storage space limitation are the following (where the index $i \in I = \{1, 2, \ldots, n\}$ denotes the products):

- $Q_i$ is the production batch size (variable)
- $K_i$ is the fixed cost per production batch (parameter)
- $D_i$ is the annual demand of the product (parameter)
- $h_i$ is the unit holding cost per year (parameter)
- $T$ is the cycle time (variable)
- $P_i$ is the annual production rate (parameter)
- $a_i$ is the storage area requirement per inventory unit (parameter)
- $A$ is the maximum available storage area (parameter)

The total cost function (TCU), including production setup costs, the inventory holding costs, and the constraint concerning the limitation on the storage area for all products are given by

$$\begin{aligned} \min \quad & TCU(Q_1, \ldots, Q_n) = \sum_{i=1}^{n} \frac{K_i D_i}{Q_i} + \sum_{i=1}^{n} \frac{h_i Q_i \rho_i}{2} \\ \text{s. t.} \quad & a_i Q_i \rho_i + \sum_{j>i} a_j \left( Q_j \rho_j - I D_j - \left( \sum_{k>j} \frac{Q_k D_j}{P_k} \right) - \left( \sum_{k \leq i} \frac{Q_k D_j}{P_k} \right) \right) \\ & + \sum_{j<i} a_j \left( Q_j \rho_j - \left( \sum_{k=j+1}^{i} \frac{Q_k D_j}{P_k} \right) \right) \leq A, \; i = 1, ..., n \end{aligned} \tag{1}$$

where $I = T - \sum_{i=1}^{n} \frac{Q_i}{P_i}$ (the idle time of the machine, we suppose it takes place in the end of a cycle, between the production of item n is finished and production

of item 1 starts) and $\rho_i = 1 - \frac{D_i}{P_i}$. Here we assume that $\sum_{i=1}^{n} \frac{D_i}{P_i} \leq 1$.The production batch size $Q_i$ can also be described with the cycle time $T$ according the formula $Q_i = TD_i$.

The storage constraint can be justified in the following way. First, we have to notice that the maximum storage requirement occurs at one of the n time points $(t_i, i = 1, , n)$ when the production of one of the n items is finished. This follows from the observation that for any $i \leq n$, if $a_i \rho_i Q_i \geq \sum_{k \neq i} \frac{a_k Q_i D_k}{P_i}$ (the required storage place for item $i$ during production is bigger than the storage that becomes available because of all the other product units sold based on the predicted demand), then we need more storage place at time point $t_(i+1)$ than $t_i$ (and the storage requirement continuously increases between the two time points); otherwise we need more storage place at time point $t_i$ than $t_(i+1)$ (with continuous decrease between the two points).

After using the $\rho_i = 1 - \frac{D_i}{P_i}$ substitution and replacing the cycle time in the constraint, we obtain the following form of the objective function:

$$TCU(Q_1, \ldots, Q_n) = \sum_{i=1}^{n} \frac{K_i D_i}{Q_i} + \sum_{i=1}^{n} \frac{h_i Q_i}{2} - \sum_{i=1}^{n} \frac{h_i Q_i D_i}{2P_i} \tag{2}$$

and the constraint

$$\begin{aligned} a_i Q_i - \frac{a_i Q_i D_i}{2P_i} + \sum_{j>i} a_j D_j \left( \sum_{k=i+1}^{j-1} \frac{Q_k}{P_k} \right) \\ + \sum_{j<i} a_j \left( Q_j - \left( \sum_{k=j}^{i} \frac{Q_k D_j}{P_k} \right) \right) \leq A, \; i = 1, ..., n \end{aligned} \tag{3}$$

To incorporate the uncertainty related to the estimation of demand as an input parameter for this model, we assume that the demand is uncertain but it is possible to describe it with a triangular fuzzy number (asymmetric). The fuzzy demand $(\tilde{D}_i)$ will then be represented as an asymmetrical triangular fuzzy number:

$$\tilde{D}_i = (D_i - \delta_i, D_i, D_i + \eta_i)$$

The Total Annual Cost in the fuzzy sense will be

$$T\tilde{C}U(Q_1, \ldots, Q_n) = \sum_{i=1}^{n} \frac{K_i \tilde{D}_i}{Q_i} + \sum_{i=1}^{n} \frac{h_i Q_i}{2} - \sum_{i=1}^{n} \frac{h_i Q_i \tilde{D}_i}{2P_i} \tag{4}$$

and the storage limitation with fuzzy demand can be written as

$$\begin{aligned} & a_i Q_i - \frac{a_i Q_i \tilde{D}_i}{2P_i} + \sum_{j>i} a_j \tilde{D}_j \left( \sum_{k=i+1}^{j-1} \frac{Q_k}{P_k} \right) \\ & + \sum_{j<i} a_j \left( Q_j - \left( \sum_{k=j}^{i} \frac{Q_k \tilde{D}_j}{P_k} \right) \right) \leq A,\ i = 1, ..., n \end{aligned} \tag{5}$$

We will employ two different approaches to find the optimal solution to this problem:

1. We calculate the signed distance for the total cost function and the constraint to obtain the defuzzified version of the model and then solve it as a crisp problem.
2. We use necessity measure to specify the required degree of fulfilment for the storage constraint and solve the problem based on this new constraint.

For the first approach, we need to calculate first the signed distance of an asymmetric triangular fuzzy number (representing the demand) from 0 as

$$\begin{aligned} d(\tilde{D}_I, \tilde{0}) &= \frac{1}{2} \int_0^1 ((\tilde{D}_i)_l(\alpha) + (\tilde{D}_i)_u(\alpha)) \mathrm{d}\alpha \\ &= \frac{1}{2} \int_0^1 [(D_i - \delta_i + \delta_i \alpha) + (D_i + \eta_i - \eta_i \alpha)] \mathrm{d}\alpha = D_i + \frac{\delta_i + \eta_i}{4} \end{aligned} \tag{6}$$

The defuzzified total cost function can be obtained as

$$\begin{aligned} TCU(Q_1, \ldots, Q_n) = & \sum_{i=1}^{n} \frac{K_i D_i}{Q_i} + \sum_{i=1}^{n} \frac{K_i(\eta_i - \delta_i)}{4Q_i} + \sum_{i=1}^{n} \frac{h_i Q_i}{2} \\ & - \sum_{i=1}^{n} \frac{h_i Q_i D_i}{2P_i} - \sum_{i=1}^{n} \frac{h_i Q_i (\eta_i - \delta_i)}{8P_i} \end{aligned} \tag{7}$$

and the defuzzified storage constraint can be written as

$$\begin{aligned} & a_i Q_i - \frac{a_i Q_i D_i}{2P_i} - \frac{a_i Q_i (\eta_i - \delta_i)}{8P_i} + \sum_{j>i} a_j D_j \left( \sum_{k=i+1}^{j-1} \frac{Q_k}{P_k} \right) \\ & + \sum_{j>i} \frac{a_j(\eta_i - \delta_i)}{4} \left( \sum_{k=i+1}^{j-1} \frac{Q_k}{P_k} \right) \\ & + \sum_{j<i} a_j \left( Q_j - \left( D_j + \frac{\eta_i - \delta_i}{4} \right) \left( \sum_{k=j}^{i} \frac{Q_k}{P_k} \right) \right) \leq A,\ i = 1, ..., n \end{aligned} \tag{8}$$

As for the second approach, we need to notice that the left hand side of the fuzzy constraint for every $i$ is a linear combination of triangular fuzzy numbers

and as a result of this, the whole expression also represents an asymmetric triangular fuzzy number for every $i$. According to this, we can define a triangular fuzzy number for every $i$ with the center

$$\begin{aligned} C_i =& a_i Q_i - \frac{a_i Q_i D_i}{2P_i} + \sum_{j>i} a_j D_j \left( \sum_{k=i+1}^{j-1} \frac{Q_k}{P_k} \right) \\ &+ \sum_{j<i} a_j \left( Q_j - D_j \left( \sum_{k=j}^{i} \frac{Q_k}{P_k} \right) \right) \end{aligned} \tag{9}$$

with left end-point of the support as

$$\begin{aligned} \vartheta_i =& a_i Q_i - \frac{a_i Q_i (D_i + \eta_i)}{2P_i} + \sum_{j>i} a_j (D_j - \delta_j) \left( \sum_{k=i+1}^{j-1} \frac{Q_k}{P_k} \right) \\ &+ \sum_{j<i} a_j \left( Q_j - (D_j + \eta_j) \left( \sum_{k=j}^{i} \frac{Q_k}{P_k} \right) \right) \end{aligned} \tag{10}$$

and right end-point of the support as

$$\begin{aligned} \nu_i =& a_i Q_i - \frac{a_i Q_i (D_i - \delta_i)}{2P_i} + \sum_{j>i} a_j (D_j + \eta_j) \left( \sum_{k=i+1}^{j-1} \frac{Q_k}{P_k} \right) \\ &+ \sum_{j<i} a_j \left( Q_j - (D_j - \delta_j) \left( \sum_{k=j}^{i} \frac{Q_k}{P_k} \right) \right) \end{aligned} \tag{11}$$

To use the possibility measure for evaluating the storage constraint, we have to define first to which extent we require the constraint to be satisfied (what should be the possibility), as $0 \leq \omega \leq 1$, and we require that for every $i$, the fuzzy number $\tilde{C}_i = (\vartheta_i, C_i, \nu_i)$ satisfies that $C_1 - (1-\omega)(C_i - \vartheta_i) \leq A$.

## 4 Example

In this section we will present a numerical example to compare three different approaches to solve the problem defined in (1). We will calculate the optimal solutions for the:

- crisp model
- fuzzy model through signed distance-based defuzzification
- chance constrained formulation.

This problem is a fictive one, even if the numbers are in the likely range of a real Finnish paper producer. The parameters of the model are described in Table 1.

The optimal solutions for the crisp and fuzzy case are given in Table 2. As we can observe, the approach using signed distance as the defuzzification approach

**Table 1.** Parameters for the example

| Product | $K_i$ | $D_i$ | $h_i$ | $a_i$ | $P_i$ | $\eta_i$ | $\delta_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 1500 | 900 | 3 | 10 | 5500 | 135 | 225 |
| 2 | 1000 | 400 | 3 | 15 | 5500 | 60 | 100 |
| 3 | 1200 | 700 | 3 | 8 | 5600 | 105 | 175 |
| 4 | 1300 | 700 | 3 | 12 | 5500 | 105 | 175 |
| 5 | 1400 | 500 | 3 | 9 | 5500 | 75 | 125 |
| 6 | 900 | 800 | 3 | 11 | 5800 | 120 | 200 |
| 7 | 1300 | 800 | 3 | 13 | 6200 | 120 | 200 |
| 8 | 1100 | 900 | 3 | 14 | 6000 | 135 | 225 |

results in a slightly lower total cost value. The optimal batch size is lower for every type of product. The possible explanation is that, since we accounted for the uncertainty and the membership functions are specified in a way that the right width is larger than the left (there is more uncertainty concerning the upper bound for demand), we need to produce less units at a time and have shorter cycle times in order to be able to react changes in the demand according to the uncertainty.

As for the chance constrained formulation, we used $\omega = 0.8$, i.i. a 80 % assurance that the available storage is enough at any given point. According to this, the total cost value decreases significantly: by accepting a specific amount of risk of running out of storage space, we can decrease the overall cost of the company. The main change in the cost is the consequence of the higher production batches and as a result the lover setup costs. As we are accept the risk related to the storage space availability, we allow for larger number of units to be produced.

**Table 2.** Optimal solutions for the example with the different approaches

| Product | Crisp model | Signed-distance approach | Chance constrained |
|---|---|---|---|
| TCU | 7969.37 | 7923.99 | 7373.93 |
| $Q_1$ | 1082.81 | 1063.67 | 1172.33 |
| $Q_2$ | 640.52 | 627.48 | 701.52 |
| $Q_3$ | 1005.00 | 981.82 | 1104.13 |
| $Q_4$ | 1134.30 | 1106.32 | 1176.15 |
| $Q_5$ | 1158.12 | 1124.61 | 1228.94 |
| $Q_6$ | 1184.33 | 1192.60 | 1224.46 |
| $Q_7$ | 1162.32 | 1132.67 | 1298.00 |
| $Q_8$ | 1068.35 | 1018.69 | 1314.65 |

To perform a simple sensitivity analysis, we considered a parameter that plays an important role in the final decision, the available storage space $A$. The results of the optimal $TCU$ values for the three considered models are listed in Table 3.

The results show that the total cost value increases for all methods as we decrease the total available storage space. Additionally, we can observe that the difference between the crisp solution and the signed-distance solution increases with the storage space, while the difference with respect to the chance constrained solution decreases.

**Table 3.** The results for different values of $A$

| $A$ | Crisp | Signed distance | Chance constrained |
|---|---|---|---|
| 3000 | 9792.47 | 9752.93 | 9611.49 |
| 3500 | 8725.48 | 8683.22 | 8567.16 |
| 4000 | 7969.37 | 7923.99 | 7827.64 |
| 4500 | 7420.31 | 7371.56 | 7291.09 |

## 5 Conclusions

Inventory optimization can have a positive impact for a company both on responsiveness and cost as well as the environment. The model introduced in this paper is a variation of the EPQ model with many items, one manufacturing machine and limited storage space, with the demand represented as a triangular fuzzy number to incorporate uncertainty in the model. This allows for taking expert opinion into account when modeling uncertainties, especially when new suppliers and/or products are introduced. The model provides an optimal solution that takes these uncertainties into account.

The first main contribution of the model presented in the paper is the formulation of the storage constraint. Although there exists previous models incorporating storage capacity in EPQ models, they are usually to restrictive as they use in many cases simply the sum of the batch sizes which is clearly an overestimation of the required storage. We provided a formula that specifies the exact storage requirement that can occur. As a second contribution, we used this formula to extend the traditional and fuzzy EPQ model with uncertain demand. We solved the fuzzy model using two different approaches: defuzzification using the signed distance measure and chance constrained programming.

Our model is particularly suitable for solving optimization problems in a process industry context. An example resembling Finnish paper industry was used to illustrate the effect of limited storage space on the different solution approaches. Future research tracks will include increasing the presented model to several machines with shared inventory space. Different defuzzification methods will be needed to be used. Finally a complete sensitivity analysis of the different models would need to be done within specific problem domains (such as the Nordic paper industry).

## References


1. Björk, K.-M.: The economic production quantity problem with a finite production rate and fuzzy cycle time. In: Proceedings of the 41th Annual Hawaii International Conference on System Sciences (HICSS 2008), pp. 68–68. IEEE (2008)
2. Björk, K.-M.: A Multi-item Fuzzy Economic Production Quantity Problem with a Finite Production Rate. Int. J. Prod. Econ. 135, 702–707 (2012)
3. Björk, K.-M., Carlsson, C.: The Outcome of Imprecise Lead Times on the Distributors. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS 2005), pp. 81–90. IEEE (2005)
4. Björk, K.-M., Carlsson, C.: The Effect of Flexible Lead Times on a Paper Producer. Int. J. Prod. Econ. 107, 139–150 (2007)
5. Carlsson, C., Fullér, R.: Soft computing and the Bullwhip effect. Econ. Comp. 2(3), 1–26 (1999)
6. Chang, H.-C.: An application of fuzzy sets theory to the EOQ model with imperfect quality items. Comput. Oper. Res. 31, 2079–2092 (2004)
7. Charnes, A., Cooper, W.W.: Chance-constrained programming. Manage. Sci. 6, 73–79 (1959)
8. Harris, F.W.: How many parts to make at once. Mag. Manage. 10, 135–136 (1913)
9. Liberatore, M.J.: The EOQ model under stochastic lead time. Oper. Res. 27, 391–396 (1979)
10. Liu, B., Iwamura, K.: Chance constrained programming with fuzzy parameters. Fuzzy Set. Syst. 94, 227–237 (1998)
11. Ouyang, L.-Y., Wu, K.-S.: A minimax distribution free procedure for mixed inventory model with variable lead time. Int. J. Prod. Econ. 56, 511–516 (1998)
12. Ouyang, L.-Y., Yao, J.-S.: A minimax distribution free procedure for mixed inventory model involving variable lead time with fuzzy demand. Comp. Oper. Res. 29, 471–487 (2002)
13. Salameh, M.K., Jaber, M.Y.: Economic production quantity model for items with imperfect quality. Int. J. Prod. Econ. 64, 59–64 (2000)
14. Sanni, S.S., Chukwu, W.I.E.: An Economic order quantity model for Items with Three-parameter Weibull distribution Deterioration, Ramp-type Demand and Shortages. Appl. Math. Model. 37, 9698–9706 (2013)
15. Taleizadeh, A.A., Nematollahi, M.: An inventory control problem for deteriorating items with back-ordering and financial considerations. Appl. Math. Model. 38, 93–109 (2014)
16. Yao, J.S., Wu, K.: Ranking fuzzy numbers based on decomposition principle and signed distance. Fuzzy Set. Syst. 116, 275–288 (2000)
17. Zadeh, L.A.: Fuzzy Sets. Inform. Control 8, 338–353 (1965)
18. Zadeh, L.A.: Outline of a new Approach to the Analysis of Complex Systems and Decision Processes. IEEE T. Syst. Man Cy. SMC-3, 28–44 (1973)
19. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Set. Syst. 1, 3–28 (1978)

# Approximate Reasoning for an Efficient, Scalable and Simple Thermal Control Enhancement

Afef Denguir[1,2], François Trousset[1], and Jacky Montmain[1]

[1] LGI2P, EMA Nîmes, France
[2] LIRMM, Université Montpellier Montpellier 2, France
{afef.denguir,francois.trousset,jacky.montmain}@mines-ales.fr

**Abstract.** In order to ensure thermal energy efficiency and follow government's thermal guidance, more flexible and efficient buildings' thermal controls are required. This paper focuses on proposing scalable, efficient and simple thermal control approach based on imprecise knowledge of buildings' specificities. Its main principle is a weak data-dependency which ensures the scalability and simplicity of our thermal enhancement approach. For this, an extended thermal qualitative model is proposed. It is based on a qualitative description of influences that actions' parameters may have on buildings' thermal performances. Our thermal qualitative model is enriched by collecting and assessing previous thermal control performances. Thus, an approximate reasoning for a *smart* thermal control becomes effective based on our extended thermal qualitative model.

**Keywords:** Qualitative modeling, approximate reasoning, *smart* thermal control, *online* learning, preference based learning.

## 1 Introduction

Since the first oil crisis in 1974, buildings thermal regulation, in France, has become stricter and harder to fulfill. Thus, highly developed thermal control techniques became mandatory in order to fulfill the government's thermal guidance and decrease buildings energy consumption. However, in spite of the big advances in thermal technologies (*e.g.,* thermostat), *smart* thermal control suffers from deployment issues (*i.e.*, deployment costs, significant settings, significant measurement, *etc.*). In fact, the uniqueness of each building complicates the design of sufficiently efficient and widely applicable thermal controls which leads to additional costs each time that the solution needs to be deployed in a different building. Therefore, *smart* thermal control related researches remain relevant and focus mainly on efficient and highly reusable aspects of thermal control approaches. This paper's work can be referenced in this latter research area and contributes to building's thermal performance enhancement. *Zero Learning Data* and *Zero Setting Parameters* challenges are, hence, considered in this paper studies. For this, we propose a new approach (THPE: THermal Process Enhancement) based on an Extended Qualitative Model (EQM) in order to bypass the

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 416–426, 2014.

complexity of quantitative modeling and the insufficiency of qualitative ones. In fact, the EQM is based on a relevant, rather than precise, thermal enhancement modeling and an approximate, rather than accurate, reasoning. These features ensure the simplicity, scalability, efficiency and longevity of our THPE. This last implements an iterative enhancement process which is described in this paper: first, a review on thermal enhancement modeling is summarized and our orientation for an extended qualitative modeling is justified. Section 3 explains the THPE's overall algorithm. Important aspects dealing with uncertainty management and decision making are then detailed. In the conclusion section, some of the THPE thermal control experimentation results are displayed, as well as, some theoretical perspectives.

## 2 Summarized Review and Related Works

Efficient thermal control can be seen as a complex system control (*i.e.,* climate, thermodynamic materials properties, thermal technologies and regulation, human behavior, *etc*.). In fact, considering most of thermal process's influence factors may lead to a significant thermal control improvement. Therefore, predictive and advanced control approaches have been proposed to ensure *smart* thermal control [1-7]. Applied to the thermal context, the predictive control considers socio-economic objectives such as minimizing energy consumption and maximizing thermal comfort [1,2]. It is based on a mathematical thermal control modeling. Therefore, the more detailed and accurate the model parameters are, the more efficient the control would be. However, mathematical model design requires expertise, as well as, detailed and precise quantitative knowledge on buildings' thermal behavior. Advanced control has been applied for 20 years in *smart* thermal control [3]. It is mainly based on Artificial Intelligence (AI) techniques and aims to provide a simple, efficient and adaptive control without requiring detailed mathematical modeling. Indeed, learning techniques are used for system modeling. Two different paradigms can be distinguished: the quantitative one (*i.e.* statistical modeling [4] and AI modeling techniques such as ANNs (Ant Neural Networks) [5], SVMs (Support vector Machines) [6], *etc.)* and the qualitative one (*i.e.,* qualitative rules and expert based modeling [7]). Quantitative control modeling requires input training data which is, usually collected through *onsite* measurements, surveys, and available documentations. Data pre-treatment and post-treatments are, hence, requested in order to improve the model efficiency. Thermal control quantitative learning is obviously a complicated task which requires important computation loads. In fact, mathematical modeling is the hardest one since it requires the biggest amount of setting data and measurements. Statistical modeling is much easier than the mathematical one however it stills not sufficient and flexible for a refined *smart* thermal control. Well learned ANN and SVM models are more appropriate to ensure a refined smart thermal control. They, however, need significant computation loads, as well as, efficient and sufficient training-data. The qualitative formalism allows reducing the complexity of thermal control modeling. It can be less data dependent compared to the quantitative ones (expert knowledge could be sufficient for the *smart* control modeling). Ambiguities and accuracy's lack may affect negatively the qualitative

modeling efficiency and longevity for a continuous enhancement purposes. Nevertheless, a qualitative thermal control modeling can be easily adapted for different thermal scales such as buildings and *smart* grids. In order to ensure an efficient, scalable and simple *smart* thermal control, we have applied well-known qualitative enhancement techniques [8-11]. These techniques were proposed a long time ago by Williams [9], Kuipers [9] and others [10,11] in order to improve qualitative modeling efficiency and reduce their ambiguities. A survey is proposed in [11]. Therefore, we propose an Extended Qualitative Model (EQM) for an efficient, scalable and simple *smart* thermal control. Time-related informations, as well as, available quantitative observations have been used in order to improve the EQM reliability and accuracy. Moreover, simplified and generalized thermal behaviors have been considered for the thermal control qualitative modeling which is, also, denoted as a substantial qualitative enhancement technique. Hence, the EQM allows the abstraction of thermal specificities while maintaining a sufficiently relevant representation for thermal enhancement purposes. The approximate reasoning (THPE: THermal Process Enhancement) based on our EQM can, thus, be generalized for various thermal scales and specificities. Furthermore, it does not require any particular setting data and important computation loads.

## 3 THPE's General Approach

Our *smart* thermal control THPE is inspired from human's increasing abilities when manipulating objects. Let us consider an amateur cyclist who is learning how to efficiently ride his new bicycle. When climbing hills, the cyclist is continually trying to adapt his riding in order to maximize his speed and minimize his effort. For this, he does not know much about his bicycle metal, tires and wheels spoke compositions and measurements. He generally does not know precisely the characteristics of his climbing paths. However, over the time, the cyclist remains able to improve his climbing performances. In fact, the more he climbs, the more his riding performances will get better. Actually, his improvement is only based on simple rules and comparisons over his previous climbing. For instance, the cyclist may know basic riding rules about his bicycle rear wheel cogs: *if climbing is hard then use a bigger cog* and *if you want to go faster then use a smaller one*. Using these simple cog's rules and considering his previous climbing observations, the cyclist displays an approximate reasoning that can be illustrated by the following statements: "*This new hill looks like a previous one that exhausted me by that time. Therefore, to make it less exhausting I should try a bigger cog for this new hill", "I once tried to climb this kind of hills but every time my performances were slow. To go faster, I should use a smaller cog this time"*. Our THPE tries to reproduce the same approximate reasoning. In fact, when we are not familiar with buildings' thermal behavior, thermal control of buildings may seem intricate. Uncertainty about how relevant a thermal control is for a given thermal situation, is then in its highest level. The same reasoning remains true for the control of any complex system. However, objective observations (*i.e.*, vaguely identified physical behavior) and subjective ones (*i.e.*, human preferences) may contribute to reduce uncertainty about thermal control. Therefore, we introduce our EQM which is used to represent simplified thermal control rules similarly as the cog's rules in the cyclist

example. It, also, defines how these thermal control rules should be applied to ensure the control enhancement for different thermal situations. The EQM design is based on influence approximations relating thermal control parameters to thermal performances. In order to extend thermal qualitative modeling, the EQM's parameters and performances display time-related informations of the thermal general behavior. The influences, among parameters and performances, are vaguely identified from thermal general behavior models. Their accuracy is constantly improving through *online* thermal quantitative observations. Similarly as the cyclist memories about his old climbing experiences, keeping track of predate thermal control, as well as, their performances allows recalling them in similar control situations. A Thermal Control Manager (TCM) has been conceived in order to maintain thermal historical data. For each thermal control attempt, the thermal situation, controls and performances are, then, stored by the TCM. This last is described by the following set $TCM = \{k = 1..n, (S^k, CMD^k, PERF^k)\}$ where $n$ is the number of previous thermal controls and $S^k$, $CMD^k$ and $PERF^k$ are, respectively, the $k^{th}$ thermal situation (*i.e.,* outdoor and indoor temperatures, *etc*.), controls and performances. To support comparison over the previous attempts and apply approximate reasoning, AI techniques have been deployed. Fig. 1 displays the THEP's algorithm describing the general approach for a *smart* thermal control based on the EQM and TCM. $S^{new}$ refers to a new thermal situation for which an efficient thermal control needs to be computed. It, mainly, involves indoor and outdoor thermal current situations, as well as, thermal setpoints that need to be reached before occupants show up. Setpoints can, also, be efficiently identified based on an overall aggregation function (*i.e.,* thermal comfort), as well as, thermal indoor and outdoor fluctuations [12,13]. In this paper, we particularly focus on THPE's aspects dealing with reducing uncertainty about buildings' thermal control. Thus, we start by explaining our different approaches used for decreasing uncertainty about the EQM influence approximations. In order to ensure an accurate thermal control, quantitative knowledge is, then, used (step 1 and 2 in Fig. 1). Section 5 deals with uncertainty about the choice of these quantitative information in order to ensure an efficient and accurate thermal control.

**THEP** ( $S^{new}$ , $TCM$ )
  if $TCM = \emptyset$ then call the energy manager else
    1. Compute $TCM^* \subseteq TCM$ where, $\forall (S, CMD, PERF) \in TCM^*$, $S$ is similar to $S^{new}$ (section 5)
    if $TCM^* = \emptyset$ then call the energy manager else
      2. Find $(S^*, CMD^*, PERF^*) \mid \forall (S, CMD, PERF) \in TCM^*$, $CMD^*$ is most favored for $S^{new}$ (section 5)
      3. Compute $CMD^{new}$ for $S^{new}$ based on the EQM (section 4) and the quantitative information of $CMD^*$
      4. Apply $CMD^{new}$ and update the $TCM$ with the new attempt $\left(S^{new}, CMD^{new}, PERF^{new}\right)$
    end if
  end if
end

**Fig. 1.** THPE general algorithm

## 4 Influence Approximations

In order to ensure the THPE weak dependency *w.r.t.* each building's thermal specificities, the EQM applies an event-based representation [14] for the thermal control laws description. This last is more relevant than a classical sampled time representation. It is, also, considered sufficient for the thermal control laws' description since steps and ramps signals are usually used for the thermal regulations. For instance, the EQM considers the thermal control starting time which is useful to improve control delays. For each thermal control law $\mathcal{L}(t)$ we associate a control parameter vector $C = (\Delta t, \Delta p, \Delta y)$. These 3 control events are described by the thermal example showed in Fig. 2 and refer, respectively, to $\mathcal{L}(t)$ delay (time-gap between $\mathcal{L}(t)$ starting time $t_1$ and thermal control starting time $t_0$), gradient (characterized by the time-gap between $\mathcal{L}(t)$ highest $y_1$ and lowest $y_0$ values) and amplitude (height-gap between $\mathcal{L}(t)$ highest and lowest values). $CMD$ refers to the set of control parameter vectors $C$ applied on all building's actuators. Rather than building's thermal profiles, thermal performances are considered in order to ensure the EQM weak dependency *w.r.t.* each building's thermal specificities. Indeed, the performance vector $P = (cost, comfort, flexibility)$ describing thermal energy consumption, stationary thermal comfort and setpoints' achievement delay, ensures building's thermal assessment in our EQM. $PERF$ corresponds, then, to the set of all building's rooms thermal performance vectors $P$. General thermal behaviors have been studied in order to identify how each control parameter influences the considered thermal performance. Tab. 1 describes, for our EQM, the gradient directions computed over each performance *w.r.t.* each control parameter. Considering gradient directions rather than precise derivative values ensures the EQM's weak dependency *w.r.t.* building's thermal specificities. Hence, the EQM's accuracy may be lacking. However combined to thermal quantitative measurements ($CMD^*$ in Fig. 1), the gradient direction based influences are considered sufficient for the THPE's thermal enhancement. For each thermal performance $j$, where $j \in S_P$ and $S_P$ is the considered thermal performance set (*e.g.*, $S_P = \{cost, comfort, flexibility\}$), and control parameter $i$, where $i \in S_C$ and $S_C$ is the considered control parameter set (*e.g.*, $S_C = \{\Delta t, \Delta p, \Delta y\}$), an influence function $F_{ij} : V_i^C \times V_j^P \rightarrow \{-, 0, +\}$ is defined, where values of thermal control parameters $c_i$, $\forall i \in S_C$, and performances $p_j$, $\forall j \in S_P$, are, respectively, defined in $V_i^C$ and $V_j^P$. $F_{ij}$ indicates whether the performance $j$ increases (+) or decreases (–) *w.r.t.* variations of the control parameter $i$. A (0) valued $F_{ij}$ function indicates that the control parameter $i$ has no influence on the performance $j$. The $F_{ij}$ qualitative gains can, thus, be considered by the EQM for buildings' thermal control enhancement. Tab. 1 displays our EQM's influence functions. Objective and subjective thermal related knowledge is introduced in order to identify influence functions:

**Table 1.** Gradient direction based influences (0 means no influence)

| $S_C$ / $S_P$ | $\Delta t$ | $\Delta p$ | $\Delta y$ |
|---|---|---|---|
| *cost* | $F_{\Delta t\ cost}(c_{\Delta t}, p_{cost})$ | – | + |
| *comfort* | 0 | 0 | $F_{\Delta y\ comfort}(c_{\Delta y}, p_{comfort})$ |
| *flexibility* | – | – | + |

Objective knowledge corresponds, mainly, to interpretable physical phenomena. These latter can be easily confirmed by studying sign variations of simplified thermal behaviors. For instance, it is commonly known that, in winter time, thermal energy consumption ($cost$) increases by increasing the command law height ($\Delta y$). This is illustrated, in Tab. 1, by a constant influence function describing a gradual rule type on $V^C_{\Delta y} \times V^P_{cost}$ such as the greater the heating step amplitude is, the greater the thermal energy consumption would be. Therefore, regardless of buildings thermal specificities, qualitative thermal influence functions can be deduced from simplified physical behaviors (*e.g.,* $F_{\Delta y\ cost}$). Buildings' special features can occasionally be responsible of $F_{ij}$'s sign variations (*e.g.,* $F_{\Delta t\ cost}$). In this case, simple learning techniques are applied over the TCM's previous attempts in order to specifically identify each building's bending points. For instance, $F_{\Delta t\ cost}$ depends on building ventilation and insulation properties: starting the heating process earlier or later impacts differently the thermal energy consumption. Fig. 3 shows some possible shapes of the continuous function relating $c_{\Delta t}$ to $p_{cost}$ measurements. The shape of this function is obtained from the simplified thermal behavior (*i.e.,* in some cases, $F_{\Delta t\ cost}$ displays a maximum. Otherwise it is decreasing for any $c_{\Delta t}$ value). The maximum remains to be identified. Fig. 3's displayed maximums can be explained by the fact that, when outdoor temperature is lower than the indoor one, building's ambient temperature decreases until the control law is launched at time $t_1$. The $c_{\Delta t}$'s interval for which $cost$ increases refers to situations where it is more costly to start heating for a short time from a low temperature than heating the building for a longer time but starting from a higher temperature. The decreasing $p_{cost}$ *w.r.t.* $c_{\Delta t}$ refers to the opposite behavior. Furthermore, the HVAC (Heating Ventilation and Air-Conditioning) system is responsible for the rapid decrease of building's ambient temperature when the heating system is off. In fact, the HVAC continuously injects a weak percentage of the outdoor air for ventilation purposes.

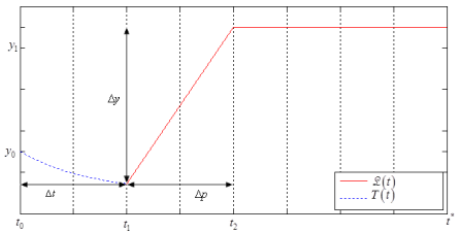



**Fig. 2.** EQM's Control laws events

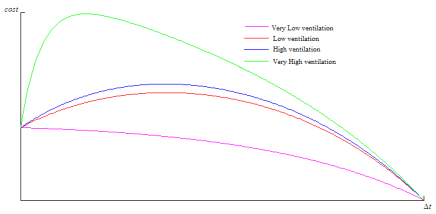



**Fig. 3.** $p_{cost}$ with regard to $c_{\Delta t}$ variations from different ventilation perspective

Consequently, we propose to use measurements in order to capture, for each building, the $c_{\Delta t} \in V_{\Delta t}^{C}$ value that entails sign variation in the continuous function (Fig. 3) and finally *online* learn $F_{\Delta t\ cost}$ function. For this, we consider the membership function $\mu_{\Delta t\ cost} : V_{\Delta t}^{C} \rightarrow [0,1]$ which describes the possibility degree that $c_{\Delta t} \in V_{\Delta t}^{C}$ may correspond to $F_{\Delta t\ cost}$'s sign variation (*i.e.,* a maximum of the continuous function relating $\Delta t$ to *cost* ). Initially, when no information is available, $\mu_{\Delta t\ cost}(c_{\Delta t})$ , $\forall c_{\Delta t} \in V_{\Delta t}^{C}$ . This case illustrates the complete ignorance regarding $F_{\Delta t\ cost}$ behavior. $\mu_{\Delta t\ cost}$ is built through *online* thermal quantitative observations. Triplets of $(c_{\Delta t}, p_{cost})$ are ranked according to $c_{\Delta t}$ . The qualitative derivative of the continuous function relating $c_{\Delta t}$ to $p_{cost}$ is, then, computed. $F_{\Delta t\ cost}$'s values can, hence, be deduced. Each new relevant thermal attempt $(S^{new}, CMD^{new}, PERF^{new})$ recommended by the THPE (section 5) and stored by the TCM, provides new triplets of $(c_{\Delta t}, p_{cost})$ which enables new $\mu_{\Delta t\ cost}$'s computations. Therefore, the ignorance interval span of $\mu_{\Delta t\ cost}$ decreases since every new qualitative derivative informs about the monotony of the continuous function. When uncertainty is not considered in the qualitative derivative computations, $\mu_{\Delta t\ cost}$ 's values belong to the $\{0,1\}$ set instead of the [0,1] interval. Uncertainty about the continuous function variations may either come from thermal disturbance or the technique used for the quantitative observations imprecision management. This kind of uncertainty management is out of this paper scope which is dedicated to discuss general uncertainty aspects in buildings' thermal control enhancement. Ideally, the *online* learning process is over when $\exists! c_{\Delta t}^{*} \in V_{\Delta t}^{C}$ such as $\mu_{\Delta t\ cost}(c_{\Delta t}^{*}) = 1$ . The membership function based *online* learning can easily be generalized in order to precisely identify more complicated buildings' thermal dependent influence functions.

Subjective knowledge can also be used in order to reduce uncertainty about buildings thermal control. This knowledge involves occupants' expectations *w.r.t.* building's performances and usages. Preference models have been, then, considered. They contribute, as well, to improve our EQM efficiency. The considered preference models

can rather be buildings dependent or independent. For instance, in Tab. 1, $F_{\Delta y\ comfort}$ influence function relating $c_{\Delta y}$ to $p_{comfort}$ measurements, is built from an overall thermal performance model that captures the multidimensional concept of *thermal comfort* [12,13]. $F_{\Delta y\ comfort}$ values can thus be identified considering building's occupants thermal sensations as well as thermal context variations (*i.e.,* humidity and sunshine characteristics). In fact, depending on the thermal context, an increasing ambient temperature may either improve or distract the occupant's thermal sensation. Hence, $F_{\Delta y\ comfort}$. acknowledge sign variations since thermal command law amplitude influence building's ambient temperature. Using thermal comfort standard such as the PPD [15] index is useful to ensure the EQM independency toward buildings' thermal properties. As the PPD formalism is complex and inadequate for control purposes, a MAUT (Multi Attribute Utility Theory) version called CIPPD has been proposed to make it easily interpretable [12,13]. The CIPPD is based on utility functions defined for each thermal comfort attribute (*i.e.,* ambient temperature, humidity, radiant temperature, air speed, *etc.*). Attributes' utilities are then aggregated to compute the comfort performance. For more information about the *thermal comfort* based control enhancement, please refer to our previous works [12,13] where you can find an extended discussion about the *thermal comfort* related issues. Considering the CIPPD's analytic form, $\mathrm{sgn}(d(u_T(T))/dT)$ function, where $T$ refers to the ambient temperature and $u_T$ to its related utility function, provides $F_{\Delta y\ comfort}$. values.

Once the EQM influences are approximated using thermal objective and subjective knowledge, thermal enhancement control can then be operated. Contradictory influences on thermal performances can, simply, be resolved by considering user's priorities. For instance, building's occupants may be more demanding about their thermal comfort. The EQM will then give priority to the *comfort* performance optimization, then *flexibility* and last the *cost* performance. Hence, based on the EQM, it becomes possible to recommend control parameters increase/decrease. Step 3 of the THPE (Fig. 1) is, then, as follows: the quantitative information $(S^*, CMD^*, PERF^*) \in TCM$, computed in step 2, provides the most favored prior attempt *w.r.t.* the current situation. Then, the EQM's rules are applied to compute a more likely *better* command law $CMD^{new}$ from $CMD^*$. The most favored $(S^*, CMD^*, PERF^*) \in TCM$ used to improve the EQM accuracy enhancement is explained in section 5.

## 5 Quantitative Knowledge Choice

The EQM's approximate reasoning is based on the selection of the quantitative control statement $(S^*, CMD^*, PERF^*) \in TCM$ as explained in Fig. 1. From one hand, $(S^*, CMD^*, PERF^*)$ is chosen such as $S^*$ is as similar as possible to $S^{new}$ (step 1 in Fig. 1), and, from the other hand, $PERF^*$ correspond to prior best realized thermal performances (step 2 in Fig. 1). Three decision criteria have been considered in order to identify the most likely favored previous attempt stored by the *TCM*: ***i.*** The first

one is similarity between previous thermal situations $S$ and the new one $S^{new}$. It allows overcoming non-linearity problems related to thermal controls (step 1 in Fig. 1) since maximizing the similarity allows linear reasoning around a setting point. Similarity between thermal situations is based on a distance $dist(S',S'')$, where $S'$ and $S''$ are two thermal situations. The smaller $dist(S',S'')$ is, the more similar $S'$ and $S''$ are. Since thermal situations are only defined by temperature measurements, there are no commensurateness problems in $dist(S',S'')$ definition. ***ii.*** The second criterion considered in TCM's statement evaluation is thermal performance. Obviously, the better the resulting thermal performances $PERF$ are, the more favored the control statement would be. For this, Multi-Criteria Decision Analyses techniques have been deployed. Thus, a preference model over the considered performances $S_P$ is identified. Firstly, utility functions ($u_{cost}$, $u_{comfort}$ and $u_{flexibility}$) are defined for each performance to ensure commensurability. They allow the assessment of each performance over the same scale which is the satisfaction degree or utility scale [0,1]. Secondly, an aggregation function is required in order to ensure the overall thermal evaluation $\mathcal{P}_r^k$ for each room $r \in R$ ($R$ corresponds to the building room's set) and prior thermal control attempt $k$. These steps are related to the energy manager preference modeling which depends on his energy policy. The preference model may be identified using indirect methods such as Macbeth. We assume that a weighted sum is sufficient to capture this preference model. When thermal control is related to a subset of rooms $R' \subseteq R$, overall thermal assessment has to consider all thermal performances over $R'$. Thus, our EQM proposes to proceed firstly by aggregating all performances from the energy consumption ($sum$), thermal comfort ($min$) and flexibility ($max$) points of view; secondly, the preference model defined for one room is applied for $R'$. We denote by $\mathcal{P}^k$ the overall building thermal assessment associated to the $k^{th}$ ($PERF^k$) prior thermal attempt stored by the TCM. ***iii.*** The last criterion considered for TCM's statement assessment is related to previous enhancement results. In fact, predate thermal controls which have led to thermal enhancement failures are disadvantaged in the future TCM's element evaluations. Therefore, we associate a set $Bad^k$ to each $(S^k, CMD^k, PERF^k) \in TCM$. $Bad^k$ gathers prior thermal controls that were computed from $(S^k, CMD^k, PERF^k)$ and led to thermal performance decreases. Considering these 3 criteria, an overall score $score^k$ (1) can be computed for each TCM's stored control in a limited neighborhood of $S^{new}$ in order to satisfy the thermal process linear behavior expected property:

$$k \in \{1,..,n\},\ score^k = \{1 - dist(S^k, S^{new})\}^{\mathcal{P}^k} * \prod_{k' \in Bad^k} \{1 - dist(S^{k'}, S^{new})\}^{\mathcal{P}^{k'}} \quad (1)$$

The favored quantitative information $(S^*, CMD^*, PERF^*) \in TCM$ used for our EQM enrichment (step 3 in Fig. 1) satisfies $score^* \geq score^k$, $\forall (S^k, CMD^k, PERF^k) \in TCM$. ($S^*$, $CMD^*$, $PERF^*$) is, then, used by the EQM in order to compute more accurate enhancement thermal control.

## 6 Conclusion and Some Experimental Results

In our previous work, we have proposed an approach allowing the computation of the most relevant target values (*i.e.*, setpoints) to be provided to the energy control system in order to improve the thermal sensation and reduce thermal energy consumption [12,13] This paper completes our approach by answering the question how to efficiently reach these setpoints without using any quantitative model and important computation loads to precisely identify each buildings thermal regulation system. Our iterative approach THPE provides thermal control recommendations, as soon as, it is deployed without needing any *a priori* learning or identification. These control recommendations are then refined thanks to quantitative observations and qualitative physical aspects related to thermal processes. Our THPE has been evaluated on a simulated building area. It ensures a quite quick and stable convergence to an *optimum* (based on the considered preference model) thermal control for every new thermal situation. In fact, a few enhancement iterations (less than 10 in most evaluation tests) are needed in order to find the *optimum* thermal control for any new thermal situation. For instance, Fig. 4 shows one room thermal enhancement process. Day 0 matches the TCM initial previous thermal control observation. Day 1 corresponds, in the same room, to the thermal profile computed for a new thermal situation based on Day 0's posteriori available quantitative information. The EQM recommendations over Day 0's control ensures 14.5% of thermal energy consumption decrease. Control enhancements are iteratively computed for the same thermal situation as Day 1 (from Day 2 to Day 5). In Fig. 4, the THPE's enhancement converges in 5 iterations where Day 5 displays the thermal profile that ensures the *optimum* thermal performances for the considered thermal situation. Our experimentations reveal about 7 to 31% for one room thermal performance enhancement and 12 to 24% for several rooms thermal enhancement. Average enhancement ensured by the THPE is evaluated to 16%. How the THPE can bypass frequent thermal control deployment issues such as quantitative data availability, it can be considered as an outstanding point compared to the existent thermal control solutions. Any comparison becomes, thus, unbalanced because of the different application conditions. Trying to operate an MPC in few days on a completely new building is not conceivable. It goes the same when asking the THPE for the same efficiency as an MPC based control. Yet, perspectives remain possible to improve our THPE efficiency. Uncertainty management in influence functions can be improved by using continuous scales membership functions. Ambiguous measurements coming from thermal disturbances (*i.e.,* windows and door opening) should complete this point. Sensors data precision can be studied as well. Qualitative interactions between the control enhancement parameters could also be studied in order to compute enhancement recommendations based on subsets of control parameters variations instead of singletons. This will warrantee the THPE's convergence to a global *optimum* rather than a local one.

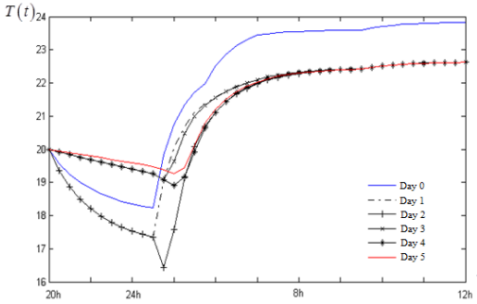



**Fig. 4.** One room thermal enhancement

## References

1. West, S.R., Ward, J.K., Wall, J.: Trial results from a model predictive control and optimisation system for commercial building HVAC. Energy and Buildings 72, 271–279 (2014)
2. Castilla, M., Álvarez, J.D., Normey-Rico, J.E., Rodríguez, F.: Thermal comfort control using a non-linear MPC strategy: A real case of study in a bioclimatic building. Journal of Process Control (2013)
3. Joint Center for Energy Management (JCEM), Final report: Artificial neural networks applied to LoanSTAR data. Technical Report TR/92/15 (1992)
4. Lei, F., Hu, P.A.: baseline model for office building energy consumption in hot summer and cold winter region. In: Proceedings of Int. Conference on Management and Service Science, vol. 6 (2009)
5. Kalogirou, S.A.: Artificial neural networks in energy applications in buildings. International Journal of Low-Carbon Technologies 3, 201–216 (2006)
6. Li, Q., Meng, Q.L., Cai, J.J., Hiroshi, Y., Akashi, M.: Applying support vector machine to predict hourly cooling load in the building. Applied Energy 86(10), 2249–2256 (2009)
7. Signh, J., Signh, N., Sharma, J.: Fuzzy modelling and control of HVAC systems - a review. Journal of scientific and Industrial Research 65, 470–476 (2006)
8. Williams, B.C.: Temporal qualitative analysis: explaining how physical systems work. In: Qualitative Reasoning about Physical Systems, pp. 133–177. Morgan Kauffmann Publishers (1989)
9. Kuipers, B.: Qualitative simulation. Artificial Intelligence 29, 289–388 (1986)
10. Dubois, D.: Order of magnitude reasoning with fuzzy relations. Artificial Intelligence 3(14), 69–94 (1989)
11. MQ&D coordinated by Dague, P.: Qualitative Reasoning: a survey of techniques and applications. AI Communications The European Journal of AI 8, 119–192 (1995)
12. Denguir, A., Trousset, F., Montmain, J.: Comfort as a Multidimensional Preference Model for Energy Efficiency Control Issues. In: Hüllermeier, E., Link, S., Fober, T., Seeger, B. (eds.) SUM 2012. LNCS, vol. 7520, pp. 486–499. Springer, Heidelberg (2012)
13. Denguir, A., Trousset, F., Montmain, J.: Research for IT Driven Energy Efficiency Based on a Multidimensional Comfort Control. In: DEXA/ICT-GLOW, pp. 18–32 (2012)
14. Montmain, J., Gentil, S.: Qualitative Event Analysis for Fault Diagnosis. In: IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, Safeprocess 1991, Baden-Baden, Allemagne (1991)
15. NF EN ISO 7730. Ergonomie des ambiances thermiques: Détermination analytique et interprétation du confort thermique à l'aide de calculs des indices PMV et PPD et du confort thermique local. AFNOR (2006)

# Robust Statistical Process Monitoring for Biological Nutrient Removal Plants

Nabila Heloulou* and Messaoud Ramdani

Laboratoire d'Automatique et Signaux de Annaba (LASA),
Universite Badji-Mokhtar de Annaba. BP. 12, Annaba 23000, Algeria
hel.nabila@yahoo.com, messaoud.ramdani@univ-annaba.dz

**Abstract.** This paper presents an approach by combining robust fuzzy principal component analysis (RFPCA) technique with the multiscale principal component analysis (MSPCA) methodology. Thus the two typical issues of industrial data, outliers and changing process conditions are solved by resulting MS-RFPCA methodology. The RFPCA is proved to be effective in mitigating the impact of noise, and MSPCA has become necessary due to the nature of complex systems in which operations occur at different scales. The efficiency of the proposed technique is illustrated on a simulated benchmark of biological nitrogen removal process.

**Keywords:** Process monitoring, robust fuzzy PCA, multiscale PCA, fault diagnosis, wavelet analysis, water treatment plant.

## 1 Introduction

The monitoring techniques allows to monitor continuously any changes in processes. For this purpose a statistical techniques have been implemented, it aims to achieve and maintain process under control. The first ideas of SPM for quality improvement go back as far as the beginning of the century. Where the principal components analysis is the most widely accepted technique to this day, the PCA technique can be seen as a projection method which allows to project the observations since space with $p$ variable dimensions towards a space with $k$ dimensions ($k \prec p$) such as a maximum of information is preserved. This fully take the nature of modern WWTPs characterized by a multitude of correlated variables [1]. Tomita et al [2] have shown the possibility of reducing the analysis from 12 variables of an activated sludge wastewater treatment down to 3 principal components which are more relevant to the process, deviation of measurements is then detected. This work has shown that the PCA is an adequate tool for representation and extracting of information. Several others recent applications of this approach and adaptation of it to wastewater treatment operations have found their way, [3], [4], [5]. Despite its success in this field, one of the most important obstacle faced is the sensitivity to outliers, also the fact that the majority of collected data from industrial processes are normally contaminated by noise makes

* N. Heloulou and M. Ramdani are with the Department of Electronics, University Badji-Mokhtar of Annaba, Algeria.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 427–436, 2014.

it unreliable in some cases. Also, in order to circumvent this difficulty, several approaches of PCA have been proposed, among the variants, robust fuzzy PCA (RF-PCA) showed promising results. In this approach, fuzzy variant of PCA uses fuzzy membership and diminish the effect of outliers by assigning small membership values to outliers in order to make it robust. Another shortcoming of conventional PCA is that modeling by PCA is done at a single scale where the actual industrial process may include events and disturbances that occur at different time-frequency range, the waste water treatment plant (WWTP) is exemplary in this respect. To solve this problem the PCA is extended for single scale to multiscale (MSPCA) modeling approach. MSPCA uses wavelet decomposition to approximately decorrelate variables autocorrelation, and also capture the linear variable correlation by PCA to extract features.

In this work, we propose the combination of MSPCA approach with a RFPCA algorithm, we thus aim to solve the problem of noise and provide a solution to the problem of monitoring during changing process conditions. The result is an effective monitoring methodology.

## 2 PCA Statistical Monitoring Principles

Principal component analysis is a multivariate statistical projection method, it is presented as a search of the subspace that maximizes the variance of the projected points, ie. the subspace that best represents the diversity of individuals through the variance-covariance structure.

The first step of this method is the construction of data matrix $X$, containing all the available data obtained by collected measurements when the process is in control. Denote the correlation matrix of $X$ as $\Sigma = X^T X/(N-1)$, and performing singular value decomposition (SVD) to the matrix $\Sigma$ yields

$$\Sigma = U \Lambda U^T$$

where $U_{n\times n}$ is a unitary matrix and $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix formed by the eigenvalues of the covariance matrix in decreasing magnitude $(\lambda_1 \geq \lambda_1 \geq \ldots \geq, \lambda_n)$. The column vectors in the matrix $U = [u_1, u_2, \ldots, u_n]$ forms a new orthogonal base of space $\mathcal{R}^n$, For dimension reduction, only the eigenvalues $\lambda_q \prec n$ are used for projection of output data in the new space. It is quite useful to consider all the directions which covers a significant portion of a total variance, these directions also be called as new basis vectors for the subspace, so the desired transformation matrix consisting of the first $q$ columns of the matrix $U$, ie. the eigenvalues belonging to the largest eigenvalues of $\Lambda$. So the first $q\,(< n)$ linear independence vectors $\hat{P} \equiv U_q = [u_1, u_2, \ldots, u_q]$ of $U$ spans the principal component subspace $\hat{S}$. The other $n-q$ vectors $\tilde{P} \equiv U_{n-q} = [u_{q+1}, u_{q+2}, \ldots, u_n]$ of $U$ spans the residual space $\tilde{S}$. The data vector $x \in \mathcal{R}^n$ can be decomposed as

$$x = \hat{x} + \tilde{x} = \hat{C}x + \tilde{C}x$$

where $\hat{x} \in \hat{S}$ and $\tilde{x} \in \tilde{S}$ are projection of $x$ on the subspaces $\hat{S}$ and $\tilde{S}$, respectively. The matrix $\hat{C} = U\hat{U}^T$ and $\tilde{C} = U\hat{U}^T$. The score vector in the space model

$t = \hat{P}^T x \in \mathcal{R}^q$ is a reduced, $q-$dimensional representation of the observed vector $x$. On the other hand, the residual $e = (I - \hat{P}\hat{P}^T)x \in \mathcal{R}^n$, represents the portion not explained by the PCA model.

In genaral, a PCA based statistical process monitoring scheme utilizes two monitoring statistics, Hotelling's $T^2$ and $Q$ statistics (SPE), which are most frequently used in the industrial processes. Typically, these indices are used to detect faults respectively in the principal component subspace $\hat{S}$ and the residual subspace $\tilde{S}$. In this work only the $Q$ statistics is used. The $SPE$ index is defined as a measure of the squared norm of the residual vector $\tilde{x}$. Box [6] shown that the confidence limit for $SPE$ from a PCA model can be calculated as :

$$\delta_\alpha^2 = \mathbf{g}\chi_{\mathbf{h},\alpha}^2 \tag{1}$$

$$\mathbf{g} = \theta_2/\theta_1 \tag{2}$$

$$\mathbf{h} = \theta_1^2/\theta_2 \tag{3}$$

where $\theta_i = \sum_{j=q+1}^{m} \lambda_j^i$ for $i = 1, 2$. Under normal operating conditions, the $SPE$ index shall be satisfied : $SPE \leq \delta_\alpha^2$.

## 3 Robust Fuzzy PCA

In actual industrial process modeling, data were often contain outliers problem. RFPCA addresses this limitation. It uses robust rules in order to replace traditional PCA and create robust fuzzy PCA. Then the influence of outliers will be reduced and consequently defects will accurately detected. The RFPCA algorithms used here were introduced in [7]. These algorithms are based on Xu and Yuille algorithms [8]. Xu and Yuille proposed an optimization function with an energy measure $e(x_i)$ subject to the membership set $u_i \in \{0, 1\}$ given as :

$$E(U, w) = \sum_{i=1}^{n} u_i e(x_i) + \eta \sum_{i=1}^{n} (1 - u_i) \tag{4}$$

The goal is to minimize $E(U, w)$ with respect to $u_i$ and $w$. Where $X = \{x_1, x_2, .., x_n\}$ is the data set, $U = \{u_i \,|i = 1, ..., n\}$ is the membership set and $\eta$ is the threshold. The variable $u_i$ serves to decide whether $x_i$ is an outlier or a sample. When $u_i = 1$ the portion of energy contributed by the sample $x_i$ is taken into consideration; otherwise $x_i$ is considered as an outlier [8]. Since $u_i$ is the binary variable and $w$ is the continuous variable, the optimization with gradient descent approach is hard to solve using gradient descent. To overcome the problem the fuzzy variant of the objective function is proposed in [7].

$$E = \sum_{i=1}^{n} u_i^m e(x_i) + \eta \sum_{i=1}^{n} (1 - u_i)^m \tag{5}$$

subject to $u_i \in [0, 1]$ and $m \in [0, 1)$. Now $u_i$ being the membership of $x_i$ belonging to data cluster and $(1 - u_i)$ is the membership of $x_i$ belonging to noise

cluster. $m$ is the so-called fuzziness variable. In this case, $e(x_i)$ measures the error between $x_i$ and the class center. This idea is similar to the C-means algorithm [9]. Since $u_i$ is now a continuous variable the difficulty of a mixture of discrete and continuous optimization can be avoided and the gradient descent approach can be used. Firstly, the gradient of equation (2) is computed respect to $u_i$ and equaled to zero, therefore :

$$u_i = \frac{1}{1 + \left(\frac{e(x_i)}{\eta}\right)^{1/((m-1))}} \tag{6}$$

Using this result in the objective function and simplifying, we obtain

$$E = \sum_{i=1}^{n} \left( \frac{1}{1 + \left(\frac{e(x_i)}{\eta}\right)^{1/((m-1))}} \right)^{m-1} e(x_i) \tag{7}$$

The gradient with respect to $w$ is

$$\frac{\delta E}{\delta w} = \beta(x_i) \left( \frac{\delta e(x_i)}{\delta w} \right) \tag{8}$$

where,

$$\beta(x_i) = \left( \frac{1}{1 + \left(\frac{e(x_i)}{\eta}\right)^{1/(m-1)}} \right)^{m} \tag{9}$$

and $m$ is the fuzziness variable. if $m = 1$, the fuzzy membership reduced to the hard membership and can be determined by following rule:

$$u_i = \begin{cases} 1 \text{ if } (e(x_i)) \langle \eta \\ 0 \text{ otherwise} \end{cases} \tag{10}$$

Now $\eta$ is a hard threshold in this situation. There is no general rule for the setting of $m$, but most papers set $m = 2$. In [7], authors derived three RFPCA algorithms, these ones are slightly different, for each algorithm the same procedure is followed except step 6 and 7. We have applied the first one in this work.

FRPCA1 algorithm :

Step 1: Initially set the iteration count $t = 1$, iteration bound $T$, learning coefficient $\alpha_0 \in (0, 1]$ soft threshold $\eta$ to a small positive value and randomly initialize the weight $w$.

Step 2: While $t$ is less than $T$, perform the next steps 3 to 9.

Step 3: Compute $\alpha_t = \alpha_0(1 - \frac{t}{T})$, set $i = 1$ and $\sigma = 0$.

Step 4: While $i$ is less than $i$, do steps 5-8.

step 5: Compute $y = w^T x_i$, $u = yw$ and $v = w^T u$.

step 6: Update the weight:

$$w^{new} = w^{old} + \alpha_T \beta\left(x_i\right)\left[y\left(x_i - u\right) + \left(y - v\right)x_i\right]$$

step 7: Update the temporary count $\delta = \delta + e_1\left(x_i\right)$.
step 8: Add 1 to $i$.
step 9: Compute $\eta = \left(\frac{\delta}{n}\right)$ and add 1 to $t$.
The weight $w$ in the updating rules converges to the principal component vector almost surely [10], [11]

## 4 Multiscale PCA Methodology

Multiscale Principal Components Analysis (MSPCA) was introduced by Bakshi (1998) that combines the merits of wavelet analysis and PCA. In particular, PCA is used to extract linear relations among variables, whereas wavelets have the ability to extract deterministic features in the measurements and decorrelate the autocorrelation among the measurements. Using wavelet, methodology involves choosing the mother wavelet from a large library of admissible functions, a selected family of wavelets with the decomposition level $L$ is applied to the signal $s$, yielding detail coefficients $D_{is(i=1\ldots s)}$ and approximation coefficient $A_{Ls}$. Next, principal component analysis is performed on each matrix of the detail scales and the matrix of approximation $\mathrm{A}_L$. The goal is to extract the correlation across the sensors. PCA control charts such as $Q$ statistics can be monitored the resulting coefficient at each scale. The wavelet coefficients of data representing normal operation are beforehand calculated, also the detection limit for $Q$ statistics are determined from them. Applying the $Q$ statistics at each scale permits identification and selection of the scales that contain the significant features representing the abnormal operation.

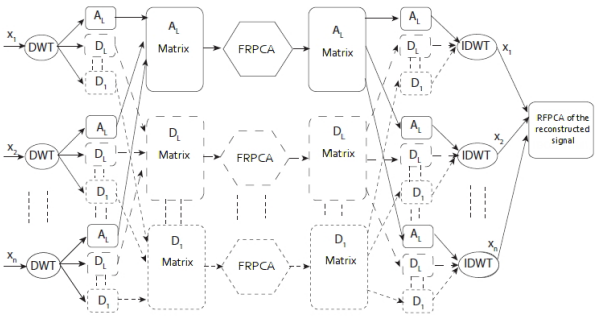



**Fig. 1.** MSPCA Methodology

## 5 MS-RFPCA Methodology

Multiscale RFPCA algorithm here intgrates RFPCA with multiscale analysis of wavelet. A decision that is crucial for the performance of the MS-RFPCA model is when the choice of the depth or number of scales of the wavelet decomposition does not take an important part of consideration. Usually, the choice of this number is an important factor in the MSPCA methodology. It should be selected to provide maximum separation between the deterministic and stochastic components of signals [12]. Therefore if we select a very small number of depth, then the last scaled signal will have a significant amount of noise that will be retained in the result of MSPCA. In the case of MSRFPCA and sight that is based on FRPCA method, thus this problem will not be asked, otherwise and when the depth is too large, the matrix of coefficients at coarser scales will have very few rows due to the dyadic down sampling, and this will affect the accuracy of the PCA at that scale. In our version (MSRFPCA), we choose the number of depth so that will not be very large to avoid the latter case. We give here the detailed algorithm including fault identification.

- Setup MSPCA Reference Model:
  1. Get the reference data when a process is under normal condition. For each variable in the reference data matrix, compute the wavelet decomposition and get the reference wavelet coefficients.
  2. For each scale, put the reference wavelet coefficients from all variables together and apply FRPCA to get the reference RFPCA model (including mean, standard deviation and PC loadings) and control limits for $T^2$ or $Q$ statistics. Repeat this procedure for all scales.
  3. Define reconstruction scenarios based on the number of decomposition level. For each reconstruction scenario, assign the selected significant scales with the corresponding reference wavelet coefficients and the insignificant scales with zeros (hard thresholding). Reconstruct the signal from the selected and thresholded coefficients for each variable. Put the reconstructed signal of all variables together and apply RFPCA to get the reference RFPCA model (including mean, standard deviation, PC loadings, and control limits) for this reconstruction scenario. Repeat the same procedure to all reconstruction scenarios.
- Online Process Monitoring:
  1. Determine the size of the moving window of dyadic length, w. Generate a data window with w samples from the real-time data by moving the time window. For each variable in the data window, compute the wavelet coefficients.
  2. For each scale, calculate $T^2$ and $Q$ scores based on the reference RFPCA model in I.(2).
  3. Compare the $T^2$ and $Q$ scores with the control limits in I.(2), retain wavelet coefficients that violate the control limits and assign those within the control limits to zero.
  4. Reconstruct the signal in the moving window variable-by-variable.

5. Since only the most recent sample (the last data in the moving window) is of interest, determine the reconstruction scenario based on the last $T^2$ and $Q$ scores of each scale and get the detection limits for the last reconstructed signal. Calculate $T^2$ and $Q$ scores of the last reconstructed signal based on the reference RFPCA model in I.(3).
6. If $T^2$ and/or $Q$ scores of the last reconstructed signal exceed the detection limits, fire an alarm.

## 6 Application on Water Treatment Plant

In this section an activated sludge model No.1 (ASM1) for nitrogen removal is presented. The basic design of this plant is shown in Fig. 2. In general, nitrogen removal proceeds in two steps. The first step is nitrification, i.e. the biological oxidation of ammonium to nitrate, this process requires an aerobic environment. During this phase ammonia nitrogen is converted into nitrite by Nitrosommas and subsequently into nitrate by Nitrobacters, for the process this phase is the crucial step. In the next step, the produced nitrate is subjected to anoxic conditions in denitrification reactor, where it is converted into harmless nitrogen gas. Anoxic/anaerobic processes operate alternately to enhance the nitrogen removal. As illustrated in Fig 2, before it enters the aeration reactor, raw wastewater $Q_{in}$ is passed by the anoxic zone, afterward the influent flow $Q_{out}$ is fed into a settler to separate the stream into the clean water and sludge, the major part of it is recycled to reactor $Q_r$, and a small part is wasted $Q_w$. The actual process model is based on the activated model sludge No.1 (ASM1) by [13]. It was adopted with two modifications: (i) the nitrification is modeled by a two step processes (the conversion of nitrite to nitrate by the nitrosoma bacteria and the conversion of nitrite to nitrate by the nitrobacters) and (ii) the hydrolysis of rapidly biodegradable substrate is included.

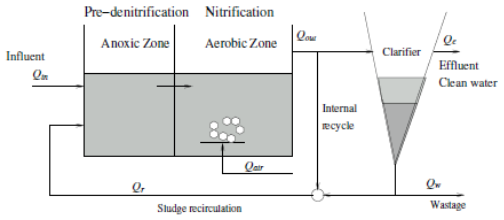



**Fig. 2.** Schematic of a typical wastewater tretment plant

Then the resulting biodegradation model consists of 18 state variables (particles and soluble concentrations) and 30 model parameters. However it is possible to reduce the model, such model is proposed by [14]. This model consists of 8 states variables : dissolved oxygen $\left(S^p_{O_2}, S^n_{O_2}\right)$, nitrate $\left(S^p_{NO_3}, S^n_{NO_3}\right)$, ammonia $\left(S^p_{NH_4}, S^n_{NH_4}\right)$, and biodegradable substrate concentrations $(S^p_S,\ S^n_S)$, for each reactor zone ($p$ and $n$ denote pre-denitrification and nitrification respectively).

So, this model consists of eight state variables :

$$x = [x_1, ..., x_8]^T = \left[S^p_{NO_3}, S^p_{O_2}, S^p_{NH_4}, S^p_S, S^n_{NO_3}, S^n_{O_2}, S^n_{NH_4}, S^n_S\right]^T.$$

More information about parameters and mathematical model can be consulted in [13].

Validation results for the developed model are shown and discussed in the next for simulated data.

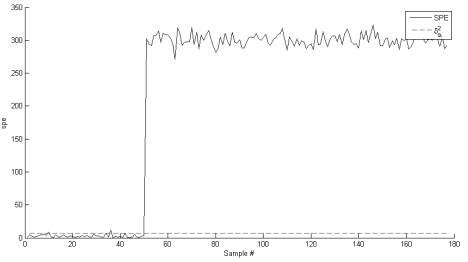



**Fig. 3.** The $SPE$ plot of linear PCA model in faulty state

Multiscale RFPCA is firstly applied to the normal operation data. Seven measured variables are constituted the measurement vector $z(k)$ available to be monitored, it is given as :
$z(k) = \left[S^p_{NO_3}(k), S^p_{O_2}(k), S^n_{NO_3}(k), S^n_{O_2}(k), u_1(k), u_2(k), u_3(k)\right]^T$
where $u_1 = Q_{in}$(influent flow rate), $u_2 = Q_r$(internal recirculation rate), $u_3 = q_{air}$ (aeration rate).

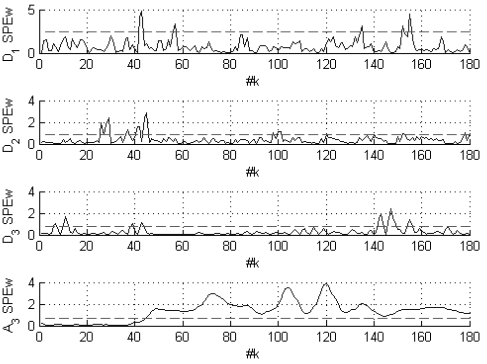



**Fig. 4.** The $SPE$ of MS-RFPCA model in faulty state

Matrix $Z$ then consists of $N$ observations of the vector $z(k)$. Also, this step includes determination of MS-FRPCA reference model : reference wavelt coefficient, reference FRPCA model, statistical limitation ($Q$).

Now, we test the fault of offset type (brusque fault) created on the level of the third sensor at at 46 sample time. Fig. 3 and Fig. 4 are the $SPE$ plot of the classical PCA and the MS-FRPCA approach respectively. According to Fig. 3, the $SPE$ plot shows the distinct change only after a delay of 5 day, whereas the multi-scale $SPE$ results (Fig. 4) shows that the $SPE$ violates 95% $SPE$ confidence limit before the $5^{th}$day. So, the MS-FRPCA can detect the ramp fault earlier than traditional PCA which ensure plant safety. Scale $Q$ charts can help to determine the nature of a disturbance. The $SPE$ plot in Fig. 4 shows a fault detected in the wavelet approximation model of MS-RFPCA, that results in change which occur in low frequencies.

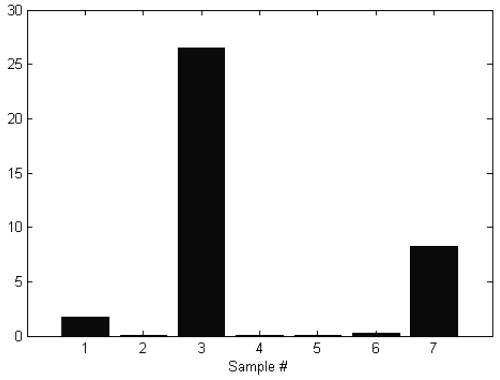



**Fig. 5.** Fault isolation using contribution to SPE (fault in the third sensor)

Our study was not only dedicated to the detection of fault to a certain level but also to the detection of faulty sensor. To identify the faulty sensor, it was exploited the contributions approach to detection index $SPE$, Fig. 5 indicates clearly that the fault sensor is $s_x = 3$.

In this work, only the sensor faults have been considered. Althought, it is important to model fault process due to different conditions, such as the toxicity shock fault caused by a reduction in the normal growth of heterotrophic organisms, there is also an inhabitation fault produced by hospital waste that can contain bacteria, another fault process that can be considered is bulking fault produced by the growth of filamentous microorganisms in the active sludge. In the other side, the method has been tested with drift fault, other different kind of sensor faults (drift, bias, precision degradation,...) can be handled by the proposed monitoring scheme. We will take these points into considerations in the next work.

## 7 Conclusions

WWTP posed an interesting challenge from the point of changing process conditions. MS-RFPCA methodology is used to monitor WWTP data during changing operational conditions. It is based firstly on time-scale decomposition in terms of increasing the sensivity of the monitoring, and secondly on RFPCA in terms of reducing noise sensivity. The results showed the advantages of the proposed monitoring method for continuous wastewater treatment plant.

## References

1. Lee, J.-M., Yoo, C., Lee, I.-B.: Statistical monitoring of dynamic processes based on dynamic independent component analysis. Chem. Engng. Sci. 59, 2995–3006 (2004)
2. Tomita, R.K., Park, S.W., Sotomayor, O.A.Z.: Analysis of activated sludge process using Multivariate statistical tools-a PCA approach. Chem. Engng. J. 90, 283–290 (2002)
3. Lee, J.-M., Yoo, C., Choi, S.W., Vanrolleghem, P.A., Lee, I.-B.: Nonlinear process monitoring using kernel principal component analysis. Chemical Eng. Sci. 59, 223–234 (2004)
4. Aguado, D., Rosen, C.: Multivariate statistical monitoring of continuous wastewater treatment plants. Engng. Applications of Artificial Intelligence 21, 1080–1091 (2008)
5. Corona, F., Mulas, M., Haimi, H., Sundell, L., Heinonen, M., Vahala, R.: Monitoring nitrate concentrations in the denitrifying post-filtration unit of a municipal wastewater treatment plant. J. Proc. Cont. 23, 158–170 (2013)
6. Box, G.E.P.: Some theorems on quadratic forms applied in the study of analysis of variance problems: Effect of inequality of variance in one-way classification. The Annals of Mathematical Statistics 25, 290–302 (1954)
7. Yang, T.N., Wang, S.D.: Robust algorithms for principal component analysis. Pattern Recognition Letters 20, 927–933 (1999)
8. Xu, L., Yuille, L.: Robust principal compoenent analysis by self-organizing rules based on statistical physics approach. IEEE Trans. Neural Networks 6(1), 131–143 (1995)
9. Oja, E.: The Nonlinear PCA Learning Rule and Signal Separation - Mathematical Analysis. Technical Report, Helsinki University of Technology (1995)
10. Oja, E.: A simplified neuron model as a principal component analyzer. J. Math. Biol. 15, 267–273 (1982)
11. Oja, E., Karhunen, J.: On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. J. Math. Analysis and Appl. 106, 69–84 (1985)
12. Yang, Q.: Model-based and data driven fault diagnosis methods with applications to process monitoring. Ph.D. thesis. Case Western Reserve University, Electrical Engineering and Computer Sciences (2004)
13. Lopez-Arenas, T., Pulis, A., Baratti, R.: On-line monitoring of a biological process for wastewater treatment. AMIDIQ 3, 51–63 (2004)
14. Gomez-Quintero, C., Queinnec, I., Babary, J.P.: A reduced nonlinear model for an activated sludge process. In: Proceeding of ADCHEM, vol. 2, pp. 1037–1402 (2000)

# Considering Psychological Conditions in a Tsunami Evacuation Simulation

Saori Iwanaga and Yoshinori Matsuura

Department of Maritime Safety Technology, Japan Coast Guard Academy,
5-1 Wakaba, Kure, Hiroshima, 737-8512, Japan
s-iwanaga@jcga.ac.jp

**Abstract.** The Great East Japan Earthquake occurred at 14:46 JST on Friday, March 11, 2011. It was the most powerful earthquake to have hit Japan and was one of the five most powerful earthquakes in the world since modern record-keeping began in 1900. The earthquake triggered an extremely destructive tsunami with waves of up to 40.5 m in height. In this paper, in preparation for a possible Nankai Trough earthquake, we created a multi-agent tsunami evacuation simulation in Kure city to analyze the evacuation process and determine solutions to any problems. More specifically, we focus on the psychological conditions of people in the disaster area. During times of emergency, it is said that people fall into a psychological condition where they fail to evacuate. Based on the simulation results, we can confirm that people under psychological conditions require more time to evacuate than those in a normal frame of mind. Thus, people need to know more about psychological conditions during times of disaster to evacuate safely and efficiently.

**Keywords:** tsunami, evacuation, multi-agent simulation, psychological conditions, Kure city.

## 1 Introduction

The Great East Japan Earthquake was a 9.0 magnitude undersea megathrust earthquake that occurred at 14:46 JST on Friday, March 11, 2011. The earthquake's epicenter was off the Pacific coast of the northeastern region of Japan. It was the most powerful earthquake to have hit Japan and one of the five most powerful earthquakes in the world since modern recordkeeping began in 1900. The earthquake triggered an extremely destructive tsunami with waves that reached 40.5 m in height in Miyako, Iwate Prefecture. In the Sendai region, waves traveled up to 10 km inland.

After the disaster, approximately 20,000 people were either dead or missing and roughly 130,000 buildings were damaged or destroyed. The earthquake and tsunami caused extensive and severe structural damage throughout northeastern Japan, including heavy damage to roads and railways as well as fires in many areas.

After the Great East Japan Earthquake, since there were many people who were unable to effectively evacuate, damage was considered to have spread. In fact, the ratio of victims in the Great East Japan Earthquake to people was lower than that in the Meiji Sanriku Earthquake, which was an 8.2 magnitude undersea megathrust earthquake that occurred off the northeast coast of Japan on June 15, 1896. It is

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 437–446, 2014.

believed that various disaster prevention methods take effect under such conditions [1]. Actually, to date, the majority of disaster control plans have utilized various disaster prevention measures that focused on aspects such as evacuation roads and extreme flooding due to dam failures. However, noting psychological conditions and behavioral characteristics during times of disasters are also important [2]. Therefore, it is necessary to examine how we can promote quick and effective evacuation behaviors under these psychological conditions.

Indeed, after the Great East Japan Earthquake, some people successfully brought an elderly person unable evacuate on his own to a shelter from the seaside in Natori [3]. But, some people failed and were victim of tsunami. In Ishinomaki (population 65,000), 2,500 people successfully evacuated to shelters. Although, 40,000 people returned to the seaside from the hillside by car to evacuate family members. Then, the resulting traffic jam brought the streets to a standstill [4].

## 2 Related Works

There have been numerous studies on tsunamis, especially in Japan, such as the simulation of tsunami hydrodynamics performed by the Japan Cabinet Office or the Japan Coast Guard. Conversely, simulations such as Ohata [5], Helbing [6], and Mas [7] have focused on crowd behavior of during times of disaster. In the latter simulations, most of researches adopt multi-agent simulation [8]. Furthermore, Katada [9] integrated simulations of tsunami hydrodynamics and evacuations.

Hirose [10] stated that psychological conditions occur during times of disaster. However, limited studies utilize a psychology model for evacuation simulation. In addition, characteristic behaviors or evacuation actions at the time of disaster have not received significant attention. According to Hirose [10], evacuation refers to the procedure of physically moving away from danger; it is a simple, classical, and effective way of escaping from impending disaster. In general, when one should evacuate to guarantee security and, especially in the case of a tsunami, evacuating the affected area in a timely manner can spell the difference between life and death. In other words, evacuating before the tsunami arrives can significantly limit the number of victims.

## 3 Psychological Conditions Operating in Times of Disaster

### 3.1 Panic

When an abnormal situation occurs, it is generally believed that panic occurs. Panic is a sudden sensation of fear that is so strong that it can dominate or prevent reason and replace logical thinking with overwhelming feelings of anxiety and frantic agitation, which are consistent with an animalistic fight-or-flight reaction.

Hirose [10] stated that panic is rarely caused and three unique psychological conditions occur at the time of disaster: normality bias, sympathy behavior, and altruistic behavior.

### 3.2 Normality Bias

When an abnormal situation occurs, normality bias arises in people in which they believe that it is not a dangerous situation and there is nothing to do. In addition, they find relief by attempting to return to everyday life. Originally, it functions to reduce the level of anxiety or fear. However, it can sometimes make those insensitive to certain risks. Examples of normality bias in the Great East Japan Earthquake included actions such as clearing a broken window panel or repairing a broken leg of cabinet, but not evacuating from the tsunami itself.

### 3.3 Sympathy Behavior

Sympathy behavior is referred to as being imitational or infectious. In this case, people want to find relief by matching his/her decisions or actions with others. When people cannot decide actions on their own, others often become the key to deciding whether to evacuate. For example, in the Great East Japan Earthquake, such behavior functioned in actions such as willingly joining traffic jams in an attempt to reach the designated shelter.

### 3.4 Altruistic Behavior

Altruistic behavior occurs in actions such as helping people despite any dangerous situations. This behavior is believed to be affected by the moral sense that every human being possesses at some level. For example, in the Great East Japan Earthquake, such behavior could be seen in actions such as evacuating an elderly person who had difficulty evacuating on his own before thinking of oneself.

## 4 Model

Ohata et al. reproduced evacuation simulation on a multi-agent simulator called KK-MAS of Kozo Keikaku Engineering Inc. in the 2003 Tokachi Oki Earthquake. In the earthquake, the tsunami alarm warned the public within 6 min. By 16 min, the tsunami waves had reached Kushiro city with the largest waves arriving 4 hours and 13 min after the earthquake. In this paper, we reproduced the Ohata' model on following Artisoc [11]. In our model, we simulated the evacuations toward the shelters and calculated the necessary time to evacuate as well as the number of evacuees for each shelter.

### 4.1 Kure City

In this study, we supposed that the Nankai Trough Earthquake occurred in the daytime on Sunday when the majority of the residents of Kure city were at home. It is estimated that the magnitude of the earthquake was 6+ and the resulting tsunami with waves of up to 4 m in height reached the city 161 min after the earthquake. In addition, we supposed that the residents living in the area forecasted to be flooded evacuated on foot [12], especially in the lower altitude areas of less than 10 m. In our model, we adopted 10 towns in the city and 30 shelters, as shown in Fig. 1 and Table 1.

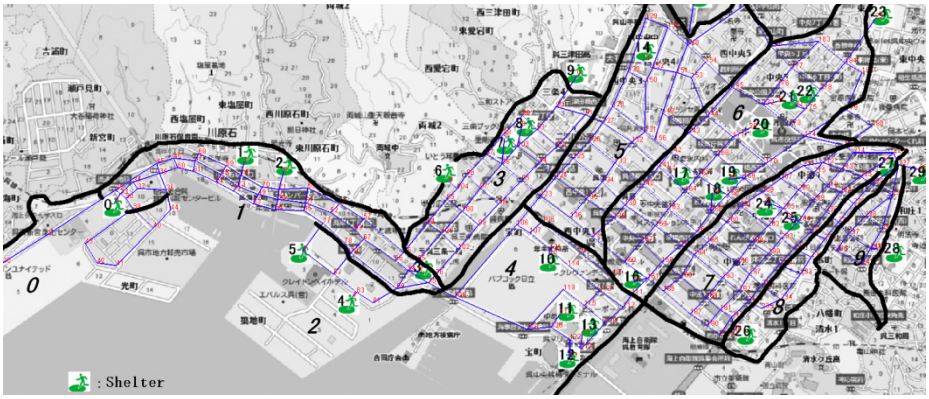



**Fig. 1.** Town, shelters and network of roads

We set the identification number of each town and population according to the basic resident register of March 31, 2013 [13]. Town 7 includes Kure city's largest mall, and the adjacent towns between Towns 5 and 8 include many residents. In this study, we set the resident agent to stand for two people, which is the average population of a household, as shown in Table 1 because the residents per household (as of March 31, 2013) averaged 1.937 in the area. According to the questionnaires in Hokkaido Nansei Oki Earthquake of 1994, many people evacuated by family unit [14].

**Table 1.** Town identification (ID), population, number of resident agents, and shelter ID

| Town ID | Population | Number of resident agents | Shelter ID | Town ID | Population | Number of resident agents | Shelter ID |
|---|---|---|---|---|---|---|---|
| 0 | 67 | 33 | 0 | 6 | 4,375 | 2,187 | 16 |
| 1 | 2,292 | 1,146 | 1 | | | | 17 |
| | | | 2 | | | | 18 |
| | | | 3 | | | | 19 |
| 2 | 64 | 32 | 4 | | | | 20 |
| | | | 5 | | | | 21 |
| 3 | 2,430 | 1,215 | 6 | | | | 22 |
| | | | 7 | | | | 23 |
| | | | 8 | 7 | 1,506 | 753 | 24 |
| | | | 9 | | | | 25 |
| 4 | 987 | 493 | 10 | 8 | 4,323 | 2,161 | 26 |
| | | | 11 | | | | 27 |
| | | | 12 | 9 | 1,173 | 586 | 28 |
| | | | 13 | | | | 29 |
| 5 | 2,960 | 1,480 | 14 | Sum | 20,177 | 10,088 | |
| | | | 15 | | | | |

It is assumed that residents evacuate to shelters located in their town. In addition, we set five shelters on our own terms since there are towns in which a shelter was not set by Kure city. Of course, for tsunami evacuations, the surrounding topography should be considered, especially in regard to hills. In this regard, there was substantial building damage on slopes in the 2001 Geiyo Earthquake [15]. Therefore, in our model, we limited the evacuations to buildings and not to hills due to possible slope failures after the earthquake.

We adopted a network type for modeling and defined the node as a node agent, which represents intersections and shelters. We also set 30 shelter nodes and 213 intersection nodes. Links represent roads and we set 384 links, as shown Fig. 1.

### 4.2 Resident Agent

Each resident agent (household) includes each of the following parameters in Table 2.

### 4.3 Behavior Rule

The resident agent is generated on a node at the first time step and goes forward to the Aim shelter. Speed is described by the distance of travel per time step. In this model, we set one time step as 1 s. The walking speed is assumed to be group speed since it has been shown that the walking speed of a group is slower because people tend to adjust their walking speeds to match slower individuals within the group. Moreover, the walking speed of a group of elderly people compared to that of younger people can be extremely different. From other studies [16], we set the Speed of a young resident agent as 0.98 m/s (standard deviation 0.20 m/s) and the Speed of an elderly resident agent as 0.84 m/s (standard deviation 0.13 m/s). According to the basic resident register, elderly people (65 years or older) make up 34% of Kure city's population.

We set the Starting time for each resident agent on the basis of the survey from the 2003 Tokachi Oki Earthquake [17], as shown in Table 3. Incidentally, according to a questionnaire and interview [10], 30%–40% of the people did not evacuate in Japan. In addition, it has been confirmed that not all people in the affected area evacuated during the Great East Japan Earthquake [17]. However, in our model, we assume that all residents evacuate the disaster-prone area.

We set the Aim shelter to be the nearest shelter in the same town. In the preliminary experiment, we compared cases in which resident agents go to the nearest shelter in the city and to the nearest shelter in the same town. We found that it is more effective when resident agents go to the nearest shelter in the same town. Based on questionnaires [17], the majority of residents evacuated to a shelter in their own towns.

Each resident agent heads for their respective Aim shelter on foot. On each node, the resident agent chooses the Aim node toward the Aim shelter. The resident agent chooses one node whose direction is closer to that of the Aim shelter. The resident agent moves forward to the Aim node from the Current node on the link. When the resident agent reaches the Aim node, the Aim node becomes the new Current node and the resident agent chooses the new Aim node until arriving at the Aim shelter.

Furthermore, we implemented the psychological condition at the time of disaster. In this paper, we treated two psychological conditions: normality bias and sympathy behavior. We will deal with altruistic behavior in another study that focuses on human psychological conditions in an evacuation simulation.

**Table 2.** Parameters of resident agent

| parameter | definition | parameter | definition |
|---|---|---|---|
| ID | ID number | Age | young or old |
| X | X coordinate | Starting time | time to start evacuation |
| Y | Y coordinate | Aim shelter | shelter to evacuate |
| Direction | direction for the aim node | Current node | node that is currently located |
| Speed | distance of travel per time step | Aim node | next node on the course to the aim shelter |
| State | on a node or on a link | | |

**Normality Bias**

As stated earlier, when an abnormal situation occurs, normality bias arises in which people believe that it is not a dangerous situation and there is nothing to do. In addition, they find relief by attempting to return to everyday life. We implemented this psychological thinking by assuming that resident agents start their evacuations late. In the model, we set 3% of the resident agents with Starting times between 900 and 2700 s after the earthquake occurred as shown in Table 3.

**Table 3.** Setting of normality bias

| Starting time[s] | Ratio [%] | |
|---|---|---|
| | Not [17] | Normality bias work |
| 0–150 | 10 | 10 |
| 150–300 | 15 | 15 |
| 300–450 | 20 | 20 |
| 450–600 | 30 | 30 |
| 600–750 | 15 | 15 |
| 750–900 | 10 | 7 |
| 900–2700 | 0 | 3 |

**Sympathy Behavior**

As previously stated, people want to find relief by matching his/her decisions or actions with others. We implement this behavior by matching the speed of a resident agent with a randomly chosen resident agent ahead of it.

### 4.4 Conditions and Evaluation

We compared six patterns by the combination of whether normality bias works and whether sympathy behavior works even if modified. We evaluated simulation results

by two evacuation values: Finish evacuation time and proportion for each shelter. We measured the Finish evacuation time in which all resident agents complete their evacuation to a shelter and we calculate the proportion of counted number of resident agents for every 30 shelters.

## 5 Simulation Results

We simulated 100 trials for each pattern and showed the average and standard deviation of the completed evacuation time in Table 4.

As shown in Table 4, when both normality bias and sympathy behavior do not work, i.e., normal conditions, it takes 1944.59 s or approximately 32 min on average to complete the evacuation. In this case, the standard deviation is 388.30 s.

On average, the Finish evacuation time extends by 6% when sympathy behavior works. In addition, the Finish evacuation time extends by 68% on average when normality bias works. That is, when psychological conditions at the time of disaster occur, it takes more time to evacuate all residents. Furthermore, we found that normality bias affect the Finish evacuation time more than sympathy behavior.

The standard deviation of the Finish evacuation time decreases when both normality bias and sympathy behavior work. That is, when some psychological conditions at the time of disaster work, the differences are small between trials and it takes 3243 s to complete the evacuation in most trials.

On the contrary, the Finish evacuation time shortens by 3% on average when modified sympathy behavior works. We believe that the reason why resident agents rush to evacuate is that they find another resident agent in a hurry. In this regard, we expect an improvement in evacuation time under special psychological conditions during times of disaster.

Fig. 2 shows the proportion of resident agents who evacuated to each shelter in same town. The horizontal axis represents the Shelter ID. When a shelter is located in town with even numbered ID, we put the Shelter ID in parentheses and painted the bars. Then, we recognize that the shelter belongs to a different town by parentheses of the Shelter ID and color of bars. We found that many resident agents evacuated to Shelter 14, which is located in Town 5 with the largest population. Shelter 15 is also located in the town, but it is inland and many resident agents near the seaside chose to concentrate on Shelter 14 instead.

**Table 4.** Finish evacuation time

| | | Normality bias | | | |
|---|---|---|---|---|---|
| | | Not | | Work | |
| | | Average | Standard deviation | Average | Standard deviation |
| Sympathy behavior | Not | 1944.59 | 388.30 | 3266.23 | 262.97 |
| | Work | 2065.18 | 255.16 | 3243.99 | 128.69 |
| | Work (modified) | 1882.82 | 317.80 | 3282.65 | 174.04 |

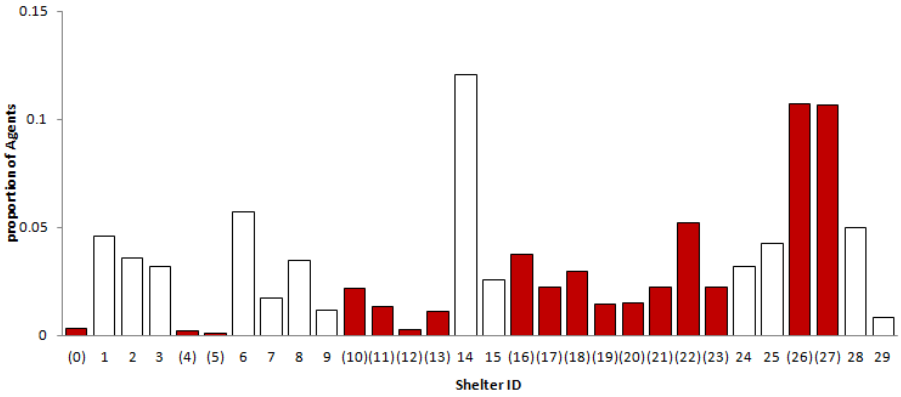



**Fig. 2.** Proportion of agents evacuated for each shelter when both Normality bias and Sympathy behavior work

We confirm through our multi-agent simulation that it takes more time to evacuate when normality bias or sympathy behavior works at the time of disaster. In addition, when two psychological conditions work simultaneously, it increases evacuation time up to 68%.

When normality bias works, the Starting times of the resident agents are 900 and 2700 s with an average of 1800 s. On the other hand, when normality bias does not work, such Starting times are between 750 and 900 s with an average of 825 s. In this case, the difference of the averages is 975 s. However, based on the simulation results, 1321.64 s (=3266.23 – 1944.59) are needed for evacuation when normality bias works at the time of disaster than when it does not work.

In Kure city, it is assumed that the tsunami arrived in 161 min or 9660 s after the earthquake occurred. In our simulation results, the resident agents could evacuate before the arrival of the tsunami. However, there remains a risk with having such ample time. For example, after the Great East Japan Earthquake in Kushiro city, once people evacuated to a shelter, majority of them returned home to help evacuate their family members before the arrival of the largest tsunami wave [17].

As stated earlier, altruistic behavior arises at the time of disaster. By this psychological condition, completing evacuations can take more time. Moreover, as seen in Ishinomaki city [1], since many people returned to the seaside by car to pick up family members, substantial traffic jams occurred and many people failed to evacuate the area.

## 6 Conclusion

In this study, we adopted a multi-agent simulation model that focused on psychological conditions during disaster. We confirm that it takes more time to complete evacuations if psychological conditions exist at the time of disaster and normality bias and sympathy behavior work. By recognizing that special psychological conditions work, it can promote early evacuations during such situations and present possibilities for improvements in evacuation processes.

The data in this model was based on information from the 2003 Tokachi Oki Earthquake. In this regard, the people of Tokachi are more used to tsunamis than those in Kure city. Therefore, it is possible that the Starting time is later than that in Tokachi and we must consider such local characteristics in the future. Furthermore, from the simulation results, there are some shelters that people concentrated on more than others. Most of these were located near the seaside; thus, it is important to promote early evacuations to other inland shelters while improving the seaside shelters due to their higher risk of flooding.

**Acknowledgement.** This work was supported by JSPS KAKENHI Grant Number 24510265.

## References

1. Ushiyama, M., Yokomaku, S.: The Great East Japan Earthquake and Disaster Information (Special Edition) Feature of human damage. Journal of Disaster Information Studies 10, 7–13 (2012) (in Japanese)
2. Hirose, H.: Wrest a living from the century of gigantic disaster. Shueisha, Japan (2011) (in Japanese)
3. NHK Special Collection Team, Mega quake- how human behaved at the time. Iwanami, Japan (2013) (in Japanese)
4. NHK Special TV Program, To the future the records of lives-big data of earthquake disaster (2013), http://www.nhk.or.jp/ (in Japanese)
5. Ohata, D., Takai, N., Kagami, H.: Spatial evaluation of tsunami refuge facilities in the central Kushiro city: Simulation of evacuation behavior from tsunami utilizing multi-agent system Part 2. Architectural Institute of Japan (612), 87–91 (2007) (in Japanese)
6. Helbing, D., Johansson, A.: Pedestrian, Crowd and Evacuation Dynamics. Encyclopedia of Complexity and Systems Science 16, 6476–6495 (2010)
7. Mas, E., Suppasri, A., Imamura, F., Koshimura, S.: Agent-based Simulation of the 2011 Great East Japan Earthquake/Tsunami Evacuation: An Integrated Model of Tsunami Inundation and Evacuation. Journal of Natural Disaster Science 34(1), 41–57 (2012)
8. Epstein, J.M., Axtell, R.L.: Growing Artificial Societies: Social Science from the Bottom Up (Complex Adaptive Systems). A Bradford Book (1996)
9. Katada, T., Kuwasawa, N., Yeh, H., Pancake, C.: Integrated simulation of tsunami hazards. EERI's Eighth U.S. National Conference on Earthquake Engineering, no.1727 (2006)
10. Hirose, H.: Why Do People Fail to Escape. Shueisha, Japan (2004) (in Japanese)
11. Yamakage, S.: Modeling and expanding artificial societies -The possibilities of an artificial society. Shosekikobo Hayama Publishing, Japan (2009)
12. Hiroshima Prefecture, Hiroshima prefecture assumption of inundated area by high tide and tsunami (2013) (in Japanese), http://www.pref.hiroshima.lg.jp/soshiki/4/tsunamisinsuisouteizu.html
13. Kure City, Shelters at the time of disaster (2012) (in Japanese), http://www.city.kure.lg.jp/kureinfo/info_hinan.pdf
14. Social Information Laboratory University of Tokyo, Society for the Study of Disaster and Information, Correspondence of residents and communication of disaster information in Hokkaido Nansei Oki Earthquake of 1993- Huge tsunami and evacuation behavior, pp. 11–18. Social Information Laboratory University of Tokyo, Japan (1994) (in Japanese)

15. Shonozaki, S., Fukuhara, Y., Masuno, K., Nishimura, M., Monzen, K., Sunamoto, F.: Damage by the, Geiyo Earthquake in Kure City: Part.2 Wooden Houses and Housing Lots. In: AIJ Tyugoku Chapter Architectural Research Meeting, vol. 25, pp. 5–8 (2002) (in Japanese)
16. Okada, K., Yoshida, K., Kashihara, S., Tsuji, M.: Human engineering of architecture and city-Mechanism of space and behavior. Kajima institute Publishing, Japan (1977) (in Japanese)
17. Ishii, Y., Kusakari, T., Minami, S., Takeuchi, S., Takahashi, A., Tomatsu, M.: A questionnaire to the people regarding the evacuation from the tsunami caused by the Great East Japan Earthquake: Part 1. For Kushiro city. In: AIJ Hokkaido Chapter Architectural Research Meeting (85) 435-440 (2012) (in Japanese)
18. Yoshii, H.: Evacuation behavior of residents in dangerous area of tsunami in Tokachi Oki Earthquake of 2003, emergency survey about Tokachi Oki Earthquake of 2003, group reconnaissance report about evacuation behavior from tsunami damage, Tokyo keizai University, Japan (2003) (in Japanese)

# $\beta$-Robust Solutions for the Fuzzy Open Shop Scheduling

Juan José Palacios[1], Inés González-Rodríguez[2], Camino R. Vela[1], and Jorge Puente Peinador[1]

[1] Department of Computer Science, University of Oviedo, Spain
{palaciosjuan,puente,crvela}@uniovi.es http://di002.edv.uniovi.es/iscop
[2] Department of Mathematics, Statistics and Computing, University of Cantabria, Spain
ines.gonzalez@unican.es

**Abstract.** We consider the open shop scheduling problem with uncertain durations modelled as fuzzy numbers. We define the concepts of necessary and possible $\beta$-robustness of schedules and set as our goal to maximise them. Additionally, we propose to assess solution robustness by means of Monte Carlo simulations. Experimental results using a genetic algorithm illustrate the proposals.

## 1 Introduction

Scheduling problems form an important body of research since the late fifties, with multiple applications in industry, finance and science [1]. In particular, the open shop scheduling problem models situations frequently appearing in testing components of electronic systems, in general repair facilities when repairs can be performed in an arbitrary order, as well as in certain medical diagnosis procedures. However, the open shop is NP-complete for a number of resources $m \geq 3$ and has a significantly large search space. Specific and efficient methods to solve it are necessary but still scarce, despite their increasing presence in the recent literature [2].

Traditionally, it has been assumed that problems are static and certain: all activities and their durations are precisely known in advance and do not change as the solution is being executed. However, for many real-world scheduling problems design variables are subject to perturbations or changes, causing optimal solutions to the original problem to be of little or no use in practice. Therefore, a common practical requirement is to obtain so-called *robust solutions*, which should still work satisfactorily when design variables change slightly, for instance, due to manufacturing tolerances.

A source of changes in scheduling problems is uncertainty in activity durations. There exists great diversity of approaches to dealing with this kind of uncertainty [3]. Perhaps the best-known is stochastic scheduling, although fuzzy sets and possibility theory provide an interesting alternative, with a tradeoff between the expressive power of probability and their associated computational

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 447–456, 2014.

complexity and knowledge demands. Indeed, fuzzy sets have been used in different manners in scheduling, ranging from representing incomplete or vague states of information to using fuzzy priority rules with linguistic qualifiers or preference modelling (cf. [4]).

The approaches to proactive robustness are several and varied. For instance, in stochastic settings, heuristic rules are used to include time buffers or slacks between activities in a baseline schedule [5]. In combinatorial optimisation, min-max or min-max regret criteria are applied to construct solutions having the best possible performance in the worst case [6], an approach already translated to the fuzzy framework [7],[8]. However, this may be deemed as too conservative when the worst case is not crucial and an overall acceptable performance is preferred. This is the basis for the $\beta$-robustness approach in stochastic scheduling [9], taking into account the subjective aspect of robustness through a target level specified by the decision maker so the goal is to maximise the likelihood that a solutions's actual performance is not worse than the target. This technique can be related to chance-constrained programming in linear optimisation, which has also been extended to fuzzy and fuzzy stochastic coefficients (cf. [10]).

The open shop problem with uncertainty constitutes a relatively new and complex research line. While there are many contributions to solve fuzzy job shop problems (we can cite, among others, [11],[12], [13] or [14]), the literature on fuzzy open shop is still scarce. Among the few existing proposals, a heuristic approach is proposed in [15] to minimise the expected makespan for an open shop problem with stochastic processing times and random breakdowns; in [16] the expected makespan of an open shop with fuzzy durations is minimised using a genetic algorithm hybridised with local search. Finally, in the framework of multiobjective approach, a possibilistic mixed-integer linear programming method is proposed in [17] for an OSP with setup times, fuzzy processing times and fuzzy due dates to minimise total weighted tardiness and total weighted completion times and in [18] a goal programming model based on lexicographic multiobjective optimisation of both makespan and due-date satisfaction is adopted and solved using a particle swarm algorithm.

In this paper, we intend to advance in the study of the fuzzy open shop problem, and in particular, in the search of robust solutions. In analogy to stochastic scheduling, we shall define the concepts of $\beta_*$-robust and $\beta^*$-robust schedules in terms of necessity and possibility, so the objective will then be to maximise such robustness. Then, we shall propose to perform an additional analysis of the obtained solutions using a Monte-Carlo simulation method based on the semantics of fuzzy schedules from [13]. Finally, we adapt the genetic algorithm from [19] and provide experimental results to illustrate our proposals.

## 2 The Fuzzy Open Shop Problem

The *open shop scheduling problem*, or *OSP* in short, consists in scheduling a set of $n$ jobs $J_1, \ldots, J_n$ to be processed on a set of $m$ physical resources or machines $M_1, \ldots, M_m$. Each job consists of $m$ tasks or operations, each requiring

the exclusive use of a different machine for its whole processing time without preemption, i.e. all operations must be processed without interruption. In total, there are $mn$ operations, $\{o_{ij}, 1 \leq i \leq n, 1 \leq j \leq m\}$. A solution to this problem is a *schedule*–an allocation of starting times for all operations– which is *feasible*, in the sense that all constraints hold, and is also optimal according to some criterion, most commonly minimising the makespan $C_{max}$, that is, the completion time of the last operation (and therefore, of the whole project).

In real-life applications, it is often the case that the exact time it takes to process a task is not known in advance. However, based on previous experience, an expert may have some knowledge (albeit uncertain) about the duration. The crudest representation of such knowledge would be a human-originated confidence interval; if some values appear to be more plausible than others, then a natural extension is a fuzzy interval or fuzzy number. The simplest model is a *triangular fuzzy number* or *TFN*, denoted $A = (a^1, a^2, a^3)$, given by an interval $[a^1, a^3]$ of possible values and a modal value $a^2 \in [a^1, a^3]$, so its membership function takes a triangular shape:

$$\mu_A(x) = \begin{cases} \frac{x-a^1}{a^2-a^1} & : a^1 \leq x \leq a^2 \\ \frac{x-a^3}{a^2-a^3} & : a^2 < x \leq a^3 \\ 0 & : x < a^1 \text{ or } a^3 < x \end{cases} \tag{1}$$

Triangular fuzzy numbers and more generally fuzzy intervals have been extensively studied in the literature (cf. [20]) and widely used in scheduling.

In the open shop, we essentially need two operations on fuzzy numbers, the sum and the maximum. For any bivariate continuous isotonic function $f$ and any two fuzzy numbers $A$ and $B$, if $A_\alpha = [\underline{a}_\alpha, \overline{a}_\alpha]$ denotes the $\alpha$-cut, the result $f(A, B)$ is a fuzzy number $F$ such that $F_\alpha = [f(\underline{a}_\alpha, \underline{b}_\alpha), f(\overline{a}_\alpha, \overline{b}_\alpha)]$, that is, computing the function is equivalent to computing it on every $\alpha$-cut. In particular, this is true for both the addition and the maximum. However, evaluating two sums or two maxima for every value $\alpha \in [0, 1]$ is cumbersome if not intractable in general. For the sake of simplicity and tractability of numerical calculations, we follow [11] and approximate the results of these operations by a linear interpolation evaluating only the operation on the three defining points of each TFN (an approach also taken, among others, in [12], [18] or [21]). The approximated sum coincides with the actual sum, so for any pair of TFNs $A$ and $Bs$:

$$A + B = (a^1 + b^1, a^2 + b^2, a^3 + b^3) \tag{2}$$

Regarding the maximum, for any two TFNs $A, B$, if $F = \max(A, B)$ denotes their maximum and $G = (\max\{a^1, b^1\}, \max\{a^2, b^2\}, \max\{a^3, b^3\})$ the approximated value, it holds that $\forall \alpha \in [0, 1]$, $\underline{f}_\alpha \leq \underline{g}_\alpha, \overline{f}_\alpha \leq \overline{g}_\alpha$. The approximated maximum $G$ is thus a TFN which artificially increases the value of the actual maximum $F$, although it maintains the support and modal value. This approximation can be trivially extended to the case of more than two TFNs.

Given a task processing order $\pi$, the schedule (starting and completion times of all tasks) may be computed as follows. For every task $x$ with processing time

$p_x$, let $S_x(\pi)$ and $C_x(\pi)$ denote respectively the starting and completion times of $x$, let $PM_x(\pi)$ and $SM_x(\pi)$ denote the predecessor and successor tasks of $x$ in the machine sequence provided by $\pi$, and let $PJ_x(\pi)$ and $SJ_x(\pi)$ denote respectively the predecessor and successor tasks of $x$ in the job sequence provided by $\pi$. Then the starting time $S_x(\pi)$ of $x$ is a TFN given by:

$$S_x(\pi) = \max(S_{PJ_x(\pi)} + p_{PJ_x(\pi)}, S_{PM_x(\pi)} + p_{PM_x(\pi)}), \tag{3}$$

Clearly, $C_x(\pi) = S_x(\pi) + p_x(\pi)$. If there is no possible confusion regarding the processing order, we may simplify notation by writing $S_x$ and $C_x$. The completion time of the last task to be processed according to $\pi$ thus calculated will be the makespan, denoted $C_{max}(\pi)$ or simply $C_{max}$. We obtain a *fuzzy schedule* in the sense that the starting and completion times of all tasks and the makespan are fuzzy intervals, interpreted as possibility distributions on the values that the times may take. However, notice that the task processing ordering $\pi$ that determines the schedule is deterministic; there is no uncertainty regarding the order in which tasks are to be processed.

## 3 Robust Schedules

The usual objective of deterministic scheduling of minimising the makespan could, in principle, be translated to the fuzzy framework as minimising the expected makespan $E[C_{max}]$. However, minimising the expected makespan may be criticised, since it reduces the information provided by a fuzzy makespan to a single value, thus loosing part of the information. Neither does it address the practical requirement of solution robustness. Therefore we propose instead to find the equivalent to what has been called in the stochastic framework $\beta$-robust schedules [9,22], schedules with a certain confidence level that the performance will be within a given threshold.

The membership function $\mu_D$ of a fuzzy duration $D$ may be interpreted as a possibility distribution on the real numbers [23,24], representing the set of more or less plausible, mutually exclusive values of a variable $y$ (in our case, the underlying uncertain duration). Since a degree of possibility can be viewed as an upper bound of a degree of probability, $\mu_D$ also encodes a whole family of probability distributions.

It is well known that for a given interval $I \subseteq \mathbb{R}$, the *possibility* and *necessity measure* that $D \in I$ are respectively given by $\Pi(D \in I) = \sup_{y \in I} \mu_D(y)$ and $N(D \in I) = \inf_{y \in I} 1 - \mu_D(y) = 1 - \sup_{y \notin I} \mu_D(x) = 1 - \Pi(D \notin I)$, so necessity and possibility are dual measures which provide lower and upper bounds for the probability that $y$ is in $I$ given the information '$y$ is $D$': $N(D \in I) \leq Pr(D \in I) \leq \Pi(D \in I)$. In particular, for $A = (a^1, a^2, a^3)$ a TFN, the *necessity* and the *possibility* that $A$ is less than a given real number $r$ are given by:

$$N(A \leq r) = \begin{cases} 0, & r \leq a^2, \\ \frac{r-a^2}{a^3-a^2}, & a^2 \leq r \leq a^3, \\ 1, & a^3 < r \end{cases} \Pi(A \leq r) = \begin{cases} 0, & r \leq a^1, \\ \frac{x-a^1}{a^2-a^1}, & a^1 \leq r \leq a^2, \\ 1, & a^2 < r \end{cases} \tag{4}$$

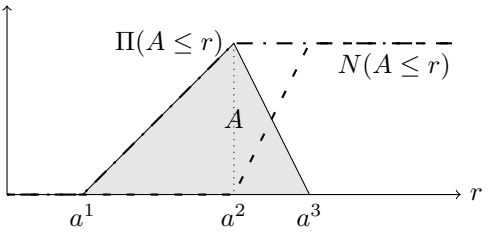



**Fig. 1.** Necessity $N(A \leq r)$ and possibility $\Pi(A \leq r)$ for varying values of $r \in \mathbb{R}$

Clearly, for any value $r$, $N(A \leq r) \leq \Pi(A \leq r)$. Figure 1 illustrates both measures.

Assuming we have a target or threshold for the makespan $C^\star$, we may want to maximise the confidence that $C_{max}$ will "for sure" be less than this threshold. In our setting, this means to maximise the necessity degree that $C_{max}$ is less than $C^\star$.

**Definition 1.** *A schedule with makespan $C_{max}$ is said to be* necessarily $\beta_*$-robust *w.r.t. a threshold $C^\star$ if and only if $\beta_* = N(C_{max} \leq C^\star)$. Analogously, the schedule is said to be* possibly $\beta^*$-robust *w.r.t. $C^\star$ iff $\beta^* = \Pi(C_{max} \leq C^\star)$. $\beta_*$ and $\beta^*$ are respectively the* degrees of necessary and possible robustness *w.r.t. the threshold $C^\star$.*

Clearly, if a schedule is $\beta^*$-robust and $\beta_*$-robust w.r.t. the same threshold, and $\beta = Pr(C_{max} \leq C^\star)$, we have that $\beta_* \leq \beta \leq \beta^*$.

The degree of necessary robustness represents the degree of confidence that the makespan will certainly be less than the threshold. In the following, we will consider that the objective will be to find a schedule maximising this confidence level, so the resulting problem may be denoted $O|fuzz\, p_i|\beta_*(C^\star)$ following the three-field notation [25]. Obviously, by maximising the degree of necessary robustness we are also maximising the possible robustness of the schedule.

## 4 Monte-Carlo Simulation Assessment

Assuming we have solved the above optimisation problem and have obtained a $\beta_*$-robust schedule w.r.t. $C^\star$, is there a means of assessing the actual robustness of such schedule? In other words, does the concept of $\beta_*$-robustness really capture the desired high-level characteristic of robustness? Here, we propose a method for an empirical assessment of solutions to the $O|fuzz\, p_i|\beta_*(C^\star)$ problem, based on using Monte-Carlo simulations and inspired by the semantics for fuzzy schedules from [13].

In [13] fuzzy schedules are interpreted as *a-priori solutions*, found when the duration of tasks is not exactly known. In this setting, it is impossible to predict what the exact time-schedule will be, because it depends on the realisation of the tasks' durations, which is not known yet. Each fuzzy schedule corresponds to

a precise ordering of tasks and it is not until tasks are executed according to this ordering that we know their real duration and, hence, know the exact schedule, the *a-posteriori solution* with exact job completion times and makespan. The practical interest of a solution to the fuzzy open shop would then lie in the ordering of tasks that it provides a priori using the available incomplete information, which should ideally yield good schedules in the moment of its practical use. Its behaviour could therefore be evaluated on a family of $K$ deterministic open shop problems, representing $K$ possible a posteriori realisations of the fuzzy problem. These may be simulated by generating an exact duration $\hat{p}_x$ for each task at random according to a probability distribution which is coherent with the fuzzy duration $p_x$.

Given a solution to the fuzzy open shop, consider the task processing order $\pi$ it provides. For a deterministic version of the problem, let $\hat{\boldsymbol{p}}$ be the matrix of precise durations, such that $\hat{p}_{ij}$, the a-posteriori duration of operation $o_{ij}$, is coherent with the constraint imposed by the fuzzy duration $p_{ij}$. The ordering $\pi$ can be used to process the operations, where the duration of each operation $o_{ij}$ is taken to be $\hat{p}_{ij}$. This yields a time-schedule with precise starting and completion times for all tasks and, in particular, a real makespan $C_{max}(\pi, \hat{\boldsymbol{p}})$, which may be under or above the threshold $C^\star$. If instead of a single deterministic instance we consider the whole family of $K$ deterministic problems, each with a duration matrix, we obtain $K$ makespan values; the proportion $\kappa$ of those values among the $K$ which are actually below the threshold $C^\star$ gives us an empirical measure of the robustness of $\pi$. If the $\beta_*$-robustness is a good measure of the schedules robustness, then a schedule with high $\beta_*$ should correspond to a high $\kappa$.

## 5 Genetic Algorithm

To solve the optimisation problem $O|fuzz\ p_i|\beta_*(C^\star)$, we propose to use the genetic algorithm (GA) from [19]. In principle, to do so it would only be necessary to substitute the fitness function therein by the $\beta_*$-robustness degree of the schedule represented by each chromosome. However, such a straightforward approach has a serious drawback: the initial population, generated at random, consists of poor schedules, with high makespan values which, most likely, will yield a value $\beta_* = 0$ for any reasonable threshold $C^\star$, thus making it impossible for the GA to evolve.

In order to overcome this drawback, we propose to adapt the GA to use an "adaptive" threshold, with successive approximations $C_0^\star > C_1^\star > \ldots$ until $C^\star$ is reached. Given the first population, a first threshold $C_0^\star$ is obtained as the most pessimistic value of the best makespan in this population, making sure that there will be chromosomes with non-zero fitness values (in fact, the individual with the best makespan will have fitness 1), thus allowing the GA to evolve. The threshold can then be updated along successive generations with new more demanding values $C_g^\star$ linearly decreasing from $C_0^\star$ to $C^\star$. This smooth updating allows the GA to evolve to robust solutions w.r.t. iteratively smaller thresholds. Finally, in order to give the GA the chance of obtaining $\beta_*$-robust solutions w.r.t.

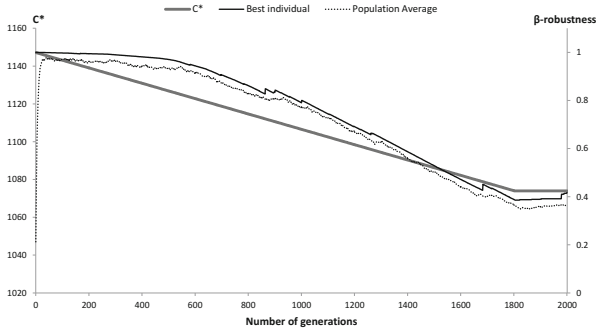



**Fig. 2.** Evolution of the best and mean solution of GA and the $C_g^\star$ values for the instance j8-per10-1 averaged across 10 runs

$C^\star$, in the last generations of the algorithm the $C^\star$ value is used to compute the $\beta_*$-robustness degree as fitness function.

## 6 Experimental Results

For the experimental study we shall use the test bed given in [16], where the authors follow [11] to generate a set of fuzzy instances from well-known open shop benchmark problems. Given a deterministic instance, each deterministic processing time $t$ is transformed into a symmetric fuzzy processing time $p(t)$ with modal value $p^2 = t$ and where values $p^1$, $p^3$ are taken as random integer values such that the resulting TFN is symmetric w.r.t. $p^2$ and its maximum range of fuzziness is 30% of $p^2$. The original benchmark consists of 6 families, denoted *J3, J4,. . . , J8*, of sizes from $3 \times 3$ to $8 \times 8$, containing 8 or 9 instances each. In this work we only consider the largest instances: i.e. the 9 instances of size $7 \times 7$ and the 8 instances of size $8 \times 8$.

In a real problem, the target value $C^\star$ would be provided by an expert with a reasonable knowledge of the problem. However, as we are using synthetic problems, such expert is not available and in consequence the target values must be set following some criterium. In our case, we have taken the best known solution $A = (a^1, a^2, a^3)$ for each instance [18] and we have defined $C^\star = a^2 + TF \times (a^3 - a^2)$, where $TF$ is a given tightness factor. To obtain the best possible performance, a parametric analysis (not reported here due to the lack of space) was conducted using $TF = 0.75$. The resulting parameter values were: population size=100, crossover probability=0.7, mutation probability=0.05, and number of generations=2000 from which the last 200 use the $C^\star$ value. The GA has been run with these parameters 10 times on each problem instance. Figure 2 shows the convergence pattern for *j8-per10-1*, one of the largest instances, with the remaining instances presenting a similar behaviour. The figure shows the evolution along 2000 generations of the fitness value of the best individual together with the mean fitness of the population and the $C_g^\star$ threshold

used at each generation $g$ to compute the $\beta_*$-robustness. As expected, we can appreciate that the algorithm's behaviour is sensitive to the $C_g^\star$ values. Initially, a less-demanding $C_0^\star$ allows the GA to evolve properly so the average quality of the population improves. After the first generations, $C_g^\star$ decreases becoming more demanding and in consequence, despite the fact that the population continues evolving, the robustness deteriorates for some generations (notice that for the same solution, its robustness value is dependent on the threshold $C^\star$). Finally, in the last iterations the goal $C^\star$ remains fixed and robustness values improve again thanks to the algorithm's evolution.

To empirically measure the robustness of the schedules obtained by the GA, we follow the Monte-Carlo simulation assessment introduced in Section 4 and generate samples of $K = 1000$ deterministic problems for each fuzzy instance, with random a-posteriori durations following a probability distribution which is coherent with the TFNs that model the fuzzy durations. We have then obtained the makespan values for each deterministic problem using the ordering provided by the GA on the fuzzy instanc, and we have finally computed the proportion $\kappa$ out of the $K$ deterministic makespan values which are below the threshold $C^\star$. Table 1 shows, for each fuzzy instance, the threshold $C^\star$, the $\beta_*$ value of the best, average and worst solution across 10 runs, the CPU time (Runtime) in seconds, and the proportion $\kappa$ obtained in the simulation for the best solution ($\kappa$-robustness). We can appreciate that even for the worst solutions $\beta_* > 0$, so in all solutions the possible $\beta^*$-robustness is 1. Moreover, the obtained "real" robustness values ($\kappa$) are always 1 or very close to 1, even in those instances

**Table 1.** Results of the GA and the *a-posteriori* analysis across the largest instances of the Brucker data set

| Instance | $C^\star$ | $\beta_*$-robustness Best | Average | Worst | Runtime | $\kappa$-robustness |
|---|---|---|---|---|---|---|
| j7-per0-0 | 1105.25 | 0.3682 | 0.2258 | 0.1082 | 9.2s. | 0.9830 |
| j7-per0-1 | 1140.00 | 0.7439 | 0.6231 | 0.4789 | 9.0s. | 1.0000 |
| j7-per0-2 | 1136.75 | 0.5493 | 0.4364 | 0.3147 | 9.0s. | 0.9980 |
| j7-per10-0 | 1099.50 | 0.7500 | 0.5294 | 0.2895 | 8.6s. | 1.0000 |
| j7-per10-1 | 1075.75 | 0.7319 | 0.5383 | 0.1972 | 8.9s. | 1.0000 |
| j7-per10-2 | 1079.75 | 0.6408 | 0.4701 | 0.2351 | 9.2s. | 1.0000 |
| j7-per20-0 | 1028.50 | 0.6477 | 0.5667 | 0.4524 | 9.0s. | 1.0000 |
| j7-per20-1 | 1075.00 | 0.7541 | 0.5041 | 0.1509 | 9.0s. | 1.0000 |
| j7-per20-2 | 1059.50 | 0.6288 | 0.3657 | 0.1508 | 9.1s. | 1.0000 |
| j8-per0-1 | 1106.50 | 0.3750 | 0.2164 | 0.0473 | 13.6s. | 0.9190 |
| j8-per0-2 | 1115.75 | 0.4696 | 0.2561 | 0.1735 | 13.8s. | 0.9630 |
| j8-per10-0 | 1110.00 | 0.9054 | 0.5723 | 0.3273 | 13.5s. | 1.0000 |
| j8-per10-1 | 1074.00 | 0.5714 | 0.4162 | 0.2692 | 13.7s. | 0.9830 |
| j8-per10-2 | 1059.25 | 0.4179 | 0.2601 | 0.0753 | 13.9s. | 0.9850 |
| j8-per20-0 | 1062.75 | 0.6433 | 0.4975 | 0.3994 | 13.6s. | 1.0000 |
| j8-per20-1 | 1048.00 | 0.7164 | 0.5445 | 0.4133 | 13.6s. | 1.0000 |
| j8-per20-2 | 1059.00 | 0.5444 | 0.4451 | 0.3299 | 13.6s. | 0.9960 |

where $\beta_*$ is smaller (e.g. j7-per0-0). This could be explained by the conservative character of the necessary robustness. In fact, in all cases where the fuzzy schedule has $\beta_* \geq 0.6$, the makespan values for all deterministic simulations are below the threshold $C^\star$.

## 7 Conclusions

We have tackled a variant of the open shop scheduling problem where uncertainty in durations is modelled using triangular fuzzy numbers. We have defined necessary and possible $\beta$-robustness in terms of scheduling and we have proposed as objective function to maximize the most pessimistic measure which is the necessary $\beta$-robustness. Moreover, we have proposed a method to empirically assess the actual robustness of the solutions. We have tested our approach using a genetic algorithm from the literature using an adaptive threshold of makespan values that overcomes the drawback of a likely random search by the GA. Based in the promising results, in the future we intend to improve on the $\beta$-robustness by adapting to the fuzzy framework the definition of $\alpha$-$\beta$-robustness, that is, for a given confidence level $\beta$ (ideally close to 1), try to minimise the threshold $\alpha$ for which this confidence is obtained (as in [22] for stochastic scheduling). We also intend to consider some kind of multiobjective approach that maximises robustness and minimises makespan.

**Acknowledgements.** This research has been supported by the Spanish Government under research grants FEDER TIN2010-20976-C02-02 and MTM2010-16051 and by the Principality of Asturias (Spain) under grant Severo Ochoa BP13106.

## References

1. Pinedo, M.L.: Scheduling. Theory, Algorithms, and Systems, 3rd edn. Springer (2008)
2. Sha, D.Y., Cheng-Yu, H.: A new particle swarm optimization for the open shop scheduling problem. Computers & Operations Research 35, 3243–3261 (2008)
3. Herroelen, W., Leus, R.: Project scheduling under uncertainty: Survey and research potentials. European Journal of Operational Research 165, 289–306 (2005)
4. Dubois, D., Fargier, H., Fortemps, P.: Fuzzy scheduling: Modelling flexible constraints vs. coping with incomplete knowledge. European Journal of Operational Research 147, 231–252 (2003)
5. Van de Vonder, S., Demeulemeester, E., Herroelen, W.: Proactive heuristic procedures for robust project scheduling: An experimental analysis. European Journal of Operational Research 189, 723–733 (2008)
6. Aissi, H., Bazgan, C., Vanderpooten, D.: Min-max and min-max regret versions of combinatorial optimization problems: A survey. European Journal of Operational Research 197, 427–438 (2009)
7. Wang, J.: A fuzzy robust scheduling approach for product development projects. European Journal of Operational Research 152, 180–194 (2004)

8. Kasperski, A., Kule, M.: Choosing robust solutions in discrete optimization problems with fuzzy costs. Fuzzy Sets and Systems 160, 667–682 (2009)
9. Daniels, R.L., Carrillo, J.E.: $\beta$-robust scheduling for single-machine systems with uncertain processing times. IIE Transactions 29, 977–985 (1997)
10. Aiche, F., Abbas, M., Dubois, D.: Chance-constrained programming with fuzzy stochastic coefficients. Fuzzy Optimization and Decision Making 12, 125–152 (2013)
11. Fortemps, P.: Jobshop scheduling with imprecise durations: a fuzzy approach. IEEE Transactions of Fuzzy Systems 7, 557–569 (1997)
12. Sakawa, M., Kubota, R.: Fuzzy programming for multiobjective job shop scheduling with fuzzy processing time and fuzzy duedate through genetic algorithms. European Journal of Operational Research 120, 393–407 (2000)
13. González Rodríguez, I., Puente, J., Vela, C.R., Varela, R.: Semantics of schedules for the fuzzy job shop problem. IEEE Transactions on Systems, Man and Cybernetics, Part A 38(3), 655–666 (2008)
14. Puente, J., Vela, C.R., González-Rodríguez, I.: Fast local search for fuzzy job shop scheduling. In: Proc. of ECAI 2010, pp. 739–744. IOS Press (2010)
15. Alcaide, D., Rodriguez-Gonzalez, A., Sicilia, J.: A heuristic approach to minimize expected makespan in open shops subject to stochastic processing times and failures. International Journal of Flexible Manufacturing Systems 17, 201–226 (2006)
16. González-Rodríguez, I., Palacios, J.J., Vela, C.R., Puente, J.: Heuristic local search for fuzzy open shop scheduling. In: Proc. FUZZ-IEEE 2010, pp. 1858–1865. IEEE (2010)
17. Noori-Darvish, S., Mahdavi, I., Mahdavi-Amiri, N.: A bi-objective possibilistic programming model for open shop scheduling problems with sequence-dependent setup times, fuzzy processing times, and fuzzy due-dates. Applied Soft Computing 12, 1399–1416 (2012)
18. Palacios, J.J., González-Rodríguez, I., Vela, C.R., Puente, J.: Swarm lexicographic goal programming for fuzzy open shop scheduling. Journal of Intelligent Manufacturing (2013)
19. Palacios, J.J., Puente, J., Vela, C.R., González-Rodríguez, I.: A genetic algorithm for the open shop problem with uncertain durations. In: Mira, J., Ferrández, J.M., Álvarez, J.R., de la Paz, F., Toledo, F.J. (eds.) IWINAC 2009, Part I. LNCS, vol. 5601, pp. 255–264. Springer, Heidelberg (2009)
20. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1986)
21. Niu, Q., Jiao, B., Gu, X.: Particle swarm optimization combined with genetic operators for job shop scheduling problem with fuzzy processing time. Applied Mathematics and Computation 205, 148–158 (2008)
22. Wu, C.W., Brown, K.N., Beck, J.C.: Scheduling with uncertain durations: Modeling $\beta$-robust scheduling with constraints. Computers & Operations Research 36, 2348–2356 (2009)
23. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York (1980)
24. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
25. Graham, R., Lawler, E., Lenstra, J., Rinnooy Kan, A.: Optimization and approximation in deterministic sequencing and scheduling: a survey. Annals of Discrete Mathematics 4, 287–326 (1979)

# Aggregation Operators on Bounded Partially Ordered Sets, Aggregative Spaces and Their Duality

Mustafa Demirci

Akdeniz University, Faculty of Sciences,
Department of Mathematics, 07058 Antalya, Turkey
demirci@akdeniz.edu.tr

**Abstract.** The present paper introduces aggregative spaces and their category **AGS**, and then establishes a dual adjunction between **AGS** and the category **Agop** of aggregation operators on bounded partially ordered sets. Spatial aggregation operators and sober aggregative spaces, enabling us to restrict the dual adjunction between **AGS** and **Agop** to a dual equivalence between the full subcategory of **Agop** consisting of spatial aggregation operators and the full subcategory of **AGS** consisting of sober aggregative spaces, will also be subjects of this paper.

**Keywords:** Aggregation operator, Aggregative space, Q-space, Generalized topological space, Category theory, Adjoint situation, Duality, Spatiality, Sobriety.

## 1 Introduction

There is a considerable interest in the studies on ($n$-ary) aggregation operators (agops for short) for replacing the particular bounded partially ordered set (poset for short) $([0,1],\leq)$ by other reasonable bounded posets, e.g. agops on $([a,b],\leq)$ in [3,10,14], agops on $(I[0,1],\preceq_w)$ (the so-called an interval-valued agops) in [12], agops on $(L^*,\leq_{L^*})$ in [8], triangular norms on a general bounded poset in [4], a general bounded lattice in [15,16], pseudo-uninorms on a general complete lattice in [17,18]. As is shown in [5,11,13], agops on general bounded posets and their category **Agop** provide a useful and an abstract framework for such studies.

The dualities between certain ordered algebraic structures and certain spaces have been an important issue in many branches of mathematics (see [6,7] and the references therein). The famous duality between the full subcategory **SobTop** of **Top** of all sober topological spaces and the full subcategory **SpatFrm** of **Frm** of all spatial frames [9] is one of such dualities. In an analogous manner to this duality, our aim in this paper is to find out an appropriate notion of space providing a categorical duality for agops. For this purpose, after the next preliminary section, we introduce the aggregative spaces and their category **AGS**, and establish a dual adjunction between **AGS** and **Agop** in Section 3. Section 4 provides the notions of spatial agops and sober aggregative spaces, and proves

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 457–465, 2014.

a dual equivalence between the full subcategory **SobAGS** of **AGS** of all sober aggregative spaces and the full subcategory **SpatAgop** of **Agop** of all spatial agops. Furthermore, the presented dual adjunction and dual equivalence have been also discussed for some full subcategories of **Agop**.

## 2 Preliminaries

### 2.1 Categorical Tools

Adjoint situations and equivalences in the category theory are essential tools for formulating the main results of this paper. By definition, an adjoint situation $(\varrho, \phi) : F \dashv G : \mathbf{C} \to \mathbf{D}$ consists of functors $G : \mathbf{C} \to \mathbf{D}$, $F : \mathbf{D} \to \mathbf{C}$, and natural transformations $id_{\mathbf{D}} \xrightarrow{\varrho} GF$ (called the unit) and $FG \xrightarrow{\phi} id_{\mathbf{C}}$ (called the co-unit) satisfying the adjunction identities $G(\phi_A) \circ \varrho_{G(A)} = id_{G(A)}$ and $\phi_{F(B)} \circ F(\varrho_B) = id_{F(B)}$ for all $A$ in $\mathbf{C}$ and $B$ in $\mathbf{D}$. If $(\varrho, \phi) : F \dashv G : \mathbf{C} \to \mathbf{D}$ is an adjoint situation for some $\varrho$ and $\phi$, then $F$ is said to be a left adjoint to $G$, $F \dashv G$ in symbols. A functor $G : \mathbf{C} \to \mathbf{D}$ is called an equivalence if it is full, faithful and isomorphism-dense. In this case, $\mathbf{C}$ and $\mathbf{D}$ are called equivalent categories, denoted by $\mathbf{C} \sim \mathbf{D}$.

**Proposition 1.** *[2,7] Given an adjoint situation $(\varrho, \phi) : F \dashv G : \mathbf{C}^{op} \to \mathbf{D}$, let $Fix(\phi)$ denote the full subcategory of $\mathbf{C}$ of all $\mathbf{C}$-objects $A$ such that $\phi_A^{op} : A \to FGA$ is a $\mathbf{C}$-isomorphism, and $Fix(\varrho)$ the full subcategory of $\mathbf{D}$ of all $\mathbf{D}$-objects $B$ such that $\varrho_B : B \to GFB$ is a $\mathbf{D}$-isomorphism. Then the following statements are true:*

*(i) The restriction of $F \dashv G$ to $[Fix(\phi)]^{op}$ and $Fix(\varrho)$ induces an equivalence $[Fix(\phi)]^{op} \sim Fix(\varrho)$.*

*(ii) If $\phi_A^{op}$ is an epimorphism in $\mathbf{C}$ for each $\mathbf{C}$-object $A$, then both $Fix(\phi)$ and $Fix(\varrho)$ are reflective in their respective categories with the reflectors $F^{op}G^{op}$ and $GF$, and reflection arrows $\phi_A^{op}$ and $\varrho_B$, resp.*

For more information about adjoint situations and equivalences, we refer to [1].

### 2.2 Aggregation Operators and Their Categories

Let $(L, \leq)$ be a bounded poset with the least element $\bot$ and the greatest element $\top$. An aggregation operator on $L$ is defined to be a function $A : \bigcup_{n \in \mathbb{N}^+} L^n \to L$ satisfying the following conditions:

(AG.1) If $\alpha_1 \leq \beta_1$, $\alpha_2 \leq \beta_2$, ... and $\alpha_n \leq \beta_n$ for all $\alpha_i$, $\beta_i \in L$, $i = 1, 2, ..., n$ $(n \in \mathbb{N}^+)$, then $A(\alpha_1, ..., \alpha_n) \leq A(\beta_1, ..., \beta_n)$.

(AG.2) $A(\alpha) = \alpha$ for all $\alpha \in L$.

(AG.3) $A(\bot, ..., \bot) = \bot$ and $A(\top, ..., \top) = \top$.

For $n \geq 2$, a function $B : L^n \to L$ is called an $n$-ary aggregation operator on $L$ iff the conditions (AG.1) and (AG.3) are satisfied. A 1-ary aggregation operator $B : L \to L$ is the identity map $id_L$ on $L$. Every aggregation operator $A$ on $L$

uniquely determines a family of $n$-ary aggregation operators $\{A_n \mid n \in \mathbb{N}^+\}$ by $A_n(\alpha_1, ..., \alpha_n) = A(\alpha_1, ..., \alpha_n)$.

With regard to the special cases of $(L, \leq)$, an aggregation operator on $L$ produces an aggregation process for fuzzy sets, interval-valued fuzzy sets, intuitionistic fuzzy sets, type 2 fuzzy sets and probabilistic metrics [5]. We further remark that whereas (AG.2) is proposed as a convention by some authors (e.g., see [3,11,13]), this condition is used to set up many interesting properties of aggregation operators such as their close connection with partially ordered groupoids in [5].

**Definition 1.** *[5] The category* ***Agop*** *of aggregation operators has as objects all triples $(L, \leq, A)$, where $(L, \leq)$ is a bounded poset and $A$ is an aggregation operator on $L$, and as morphisms all $(L, \leq, A) \xrightarrow{u} (M, \leq, B)$, where $u : (L, \leq) \to (M, \leq)$ is an order-preserving function such that $u(\bot) = \bot$, $u(\top) = \top$ and the following diagram commutes for all $n \in \mathbb{N}^+$:*

$$\begin{array}{ccc} L^n & \xrightarrow{u^n} & M^n \\ {\scriptstyle A_n}\downarrow & & \downarrow{\scriptstyle B_n} \\ L & \xrightarrow{u} & M, \end{array}$$

*i.e. $u(A(\alpha_1, ..., \alpha_n)) = B(u(\alpha_1), ..., u(\alpha_n))$ for all $\alpha_1, ..., \alpha_n \in L$. Composition and identities in* ***Agop*** *are taken from the category* ***Set*** *of sets and functions.*

**Definition 2.** *[5] (i)* ***Asagop*** *is the full subcategory of* ***Agop*** *of all $(L, \leq, A)$ such that $A$ is associative, i.e.*

$$A(\alpha_1, ..., \alpha_k, ...\alpha_n) = A_2(A_k(\alpha_1, ..., \alpha_k), A_{n-k}(\alpha_{k+1}, ..., \alpha_n))$$

*for all $n \geq 2$, $k = 1, .., n-1$ and $\alpha_1, ..., \alpha_n \in L$.*

*(ii)* ***Smasagop*** *is the full subcategory of* ***Asagop*** *of all $(L, \leq, A)$ such that $A$ is symmetric, i.e.*

$$A(\alpha_1, ..., \alpha_n) = A(\alpha_{\pi(1)}, ..., \alpha_{\pi(n)})$$

*for all $n \in \mathbb{N}^+$, $\alpha_1, ..., \alpha_n \in L$ and for all permutations $\pi(1)$,...,$\pi(n)$ of $\{1, .., n\}$.*

## 3 Aggregative Spaces and Their Relations with Aggregation Operators

### 3.1 Definition of Aggregative Spaces and Their Category

For a given set $X$, we call a subset $\tau$ of the power set $\mathcal{P}(X)$ of $X$ an aggregative system on $X$ if $\emptyset \in \tau$, $X \in \tau$, and $G_1, G_2 \in \tau$ implies $G_1 \cap G_2 \in \tau$ for all $G_1, G_2 \in \mathcal{P}(X)$. By an aggregative space, we mean a pair $(X, \tau)$ of a set $X$ and an aggregative system $\tau$ on $X$. To formulate the category of aggregative spaces, we need to recall that every function $f : X \to Y$ determines a function $f^{\leftarrow} : \mathcal{P}(Y) \to \mathcal{P}(X)$, sending each subset $G$ of $Y$ to the preimage of $G$ under $f$.

**Definition 3.** *The category of aggregative spaces, denoted by* ***AGS****, is a category whose objects are aggregative spaces, and whose morphisms* $f : (X, \tau) \to (Y, \nu)$ *are functions* $f : X \to Y$ *having the property that for every* $G \in \nu$, $f^{\leftarrow}(G) \in \tau$. *Composition and identities in* ***AGS*** *are the same as those in* ***Set****.*

It is remarkable to mention that **AGS** is a special kind of the category $(\mathcal{Z}_1, \mathcal{Z}_2)\mathbf{S}$ of $(\mathcal{Z}_1, \mathcal{Z}_2)$-spaces, which has been developed in [6]. More clearly, **AGS** is exactly the same as the category $(\mathcal{V}_\perp, \mathcal{F})\mathbf{S}$, where $\mathcal{V}_\perp$ and $\mathcal{F}$ are the subset systems, defined by $\mathcal{V}_\perp(P) = \{\emptyset\}$ and $\mathcal{F}(P) = \{G \subseteq P \mid G \text{ is finite}\}$ for every poset $P$ in [6].

### 3.2 Dual Adjunction between AGS and Agop

Our aim in this section is to show that there exists an adjoint situation $(\eta, \varepsilon) : \Omega_{AG} \dashv Pt_{AG} : \mathbf{Agop}^{op} \to \mathbf{AGS}$. For this purpose, we first establish the functors $\Omega_{AG} : \mathbf{AGS} \to \mathbf{Agop}^{op}$and $Pt_{AG} : \mathbf{Agop}^{op} \to \mathbf{AGS}$.

Every aggregative space $(X, \tau)$ induces an **Agop**-object $(\tau, \subseteq, X(\cap)_\tau)$, where $X(\cap)$ is the aggregation operator on $\mathcal{P}(X)$, defined by $X(\cap)(G_1, ..., G_n) = G_1 \cap ... \cap G_n$ and $X(\cap)(G) = G$ for every $G, G_1, ..., G_n \in \mathcal{P}(X)$ $(n \geq 2)$, and $X(\cap)_\tau$ is the restriction of $X(\cap)$ to the set $\bigcup_{n \in \mathbb{N}^+} \tau^n$. This means that the assignment of the **Agop**-object $(\tau, \subseteq, X(\cap)_\tau)$ to every aggregative space $(X, \tau)$ is a function $\Omega_{AG}$ from the objects of **AGS** to the objects of **Agop**. On the other hand, for a given **AGS**-morphism $f : (X, \tau) \to (Y, \nu)$, the restriction $f^{\leftarrow}_{|_\nu}$ of $f^{\leftarrow}$ to $\nu$ is an **Agop**-morphism $f^{\leftarrow}_{|_\nu} : (\nu, \subseteq, Y(\cap)_\nu) \to (\tau, \subseteq, X(\cap)_\tau)$. Thus, $\Omega_{AG}$ can be extended to a functor from **AGS** to $\mathbf{Agop}^{op}$:

**Proposition 2.** *The map* $\Omega_{AG} : \mathbf{AGS} \to \mathbf{Agop}^{op}$, *defined by*

$$\Omega_{AG}(X, \tau) = (\tau, \subseteq, X(\cap)_\tau) \text{ and } \Omega_{AG}(f) = \left(f^{\leftarrow}_{|_\nu}\right)^{op},$$

*is a functor.*

In the formulation of the functor $Pt_{AG} : \mathbf{Agop}^{op} \to \mathbf{AGS}$, we will use the notion of filter defined as follows.

**Definition 4.** *Let* $(L, \leq, A)$ *be an object of* ***Agop****. A subset* $G$ *of* $L$ *is called a filter of* $(L, \leq, A)$ *iff* $G$ *satisfies the next conditions:*

*(F1)* $G$ *is an upper set of* $(L, \leq)$, *i.e. for all* $\alpha, \beta \in L$, $\alpha \in G$ *and* $\alpha \leq \beta$ *imply* $\beta \in G$,

*(F2)* $\perp \notin G$,

*(F3)* $\top \in G$,

*(F4) For all* $\alpha_1, ..., \alpha_n \in L$, $\alpha_1, ..., \alpha_n \in G$ *iff* $A(\alpha_1, ..., \alpha_n) \in G$.

**Lemma 1.** *Given an* ***Agop****-object* $(L, \leq, A)$, *let* $\mathfrak{F}(L)$ *denote the set of all filters of* $(L, \leq, A)$. *For each* $a \in L$, *let* $\Psi_a = \{G \in \mathfrak{F}(L) \mid a \in G\}$ *and* $\Psi(L) = \{\Psi_a \mid a \in L\}$. *Then,* $Pt_{AG}(L, \leq, A) = (\mathfrak{F}(L), \Psi(L))$ *is an aggregative space.*

*Proof.* For each $G \in \mathfrak{F}(L)$, by (F2) and (F3) in Definition 4, $\mathfrak{F}(L) = \Psi_{\top} \in \Psi(L)$ and $\emptyset = \Psi_{\perp} \in \Psi(L)$. Furthermore, we obtain from (F4) that for all $a, b \in L$, $\Psi_a \cap \Psi_b = \Psi_{A(a,b)}$. Therefore, $\Psi(L)$ is an aggregative system on $\mathfrak{F}(L)$.

**Proposition 3.** *The map* $Pt_{AG} : \boldsymbol{Agop}^{op} \to \boldsymbol{AGS}$*, defined by*

$$Pt_{AG}\left((L, \leq, A) \overset{u}{\to} (M, \leq, B)\right) = Pt_{AG}(L, \leq, A) \overset{Pt_{AG}(u)}{\to} Pt_{AG}(M, \leq, B),$$

*where* $[Pt_{AG}(u)](G) = (u^{op})^{\leftarrow}(G)$ *for all* $G \in \mathfrak{F}(L)$*, is a functor.*

*Proof.* Lemma 1 shows that $Pt_{AG}$ maps the objects of $\mathbf{Agop}^{op}$ to the objects of **AGS**. Let $(L, \leq, A) \overset{u}{\to} (M, \leq, B)$ be an $\mathbf{Agop}^{op}$-morphism, i.e. $(M, \leq, B) \overset{u^{op}}{\to} (L, \leq, A)$ is an **Agop**-morphism. For every $G \in \mathfrak{F}(L)$, since $(u^{op})^{\leftarrow}(G) \in \mathfrak{F}(M)$, $Pt_{AG}(u) : \mathfrak{F}(L) \to \mathfrak{F}(M)$ is a set map. In addition to this, we easily see that for every $b \in M$, $[Pt_{AG}(u)]^{\leftarrow}(\Psi_b) = \Psi_{u^{op}(b)}$, i.e. $[Pt_{AG}(u)]^{\leftarrow}(V) \in \Psi(L)$ for every $V \in \Psi(M)$. This proves that $Pt_{AG}(u) : Pt_{AG}(L, \leq, A) \to Pt_{AG}(M, \leq, B)$ is an **AGS**-morphism. Hence, the assertion follows from the fact that $Pt_{AG}$ preserves composition and identities.

To accomplish our task in this section, we now consider two natural transformations-the unit and co-unit of the asked adjunction-given in the next two lemmas.

**Lemma 2.** *For every* ***AGS****-object* $(X, \tau)$*, the map* $\eta_{(X,\tau)} : X \to \mathfrak{F}(\tau)$*, defined by* $\eta_{(X,\tau)}(x) = \tau(x) = \{G \in \tau \mid x \in G\}$*, is an* ***AGS****-morphism* $(X, \tau) \to Pt_{AG}\Omega_{AG}(X, \tau)$*. Moreover,* $\eta = \left(\eta_{(X,\tau)}\right)_{(X,\tau) \in Ob(\mathbf{AGS})} : id_{\mathbf{AGS}} \to Pt_{AG}\Omega_{AG}$ *is a natural transformation.*

*Proof.* It is obvious that for every $x \in X$, $\tau(x) \in \mathfrak{F}(\tau)$, and so $\eta_{(X,\tau)} : X \to \mathfrak{F}(\tau)$ is indeed a map. To see that $\eta_{(X,\tau)} : (X, \tau) \to Pt_{AG}\Omega_{AG}(X, \tau)$ is an **AGS**-morphism, note first that

$$Pt_{AG}\Omega_{AG}(X, \tau) = (\mathfrak{F}(\tau), \Psi(\tau)).$$

Then, since $\eta^{\leftarrow}_{(X,\tau)}(\Psi_G) = G$ for every $G \in \tau$, we obtain that $\eta^{\leftarrow}_{(X,\tau)}(V) \in \tau$ for every $V \in \Psi(\tau)$, i.e. $\eta_{(X,\tau)} : (X, \tau) \to Pt_{AG}\Omega_{AG}(X, \tau)$ is an **AGS**-morphism. The proof of the second part requires only the naturality of $\eta$ which means the commutativity of the rectangle

$$\begin{array}{ccc} (X, \tau) & \overset{\eta_{(X,\tau)}}{\longrightarrow} & Pt_{AG}\Omega_{AG}(X, \tau) \\ f \downarrow & & \downarrow \quad Pt_{AG}\Omega_{AG}(f) \\ (Y, \nu) & \overset{\eta_{(Y,\nu)}}{\longrightarrow} & Pt_{AG}\Omega_{AG}(Y, \nu) \end{array}$$

for every **AGS**-morphism $(X, \tau) \overset{f}{\to} (Y, \nu)$. Since

$$[Pt_{AG}\Omega_{AG}(f)]\left(\eta_{(X,\tau)}(x)\right) = \left(f^{\leftarrow}_{|\nu}\right)^{\leftarrow}\left(\eta_{(X,\tau)}(x)\right),$$

the commutativity of the rectangle above follows from the following observation: For every $x \in X$ and every $H \in \nu$,

$$\begin{aligned} H \in \eta_{(Y,\nu)}(f(x)) &\Leftrightarrow H \in \nu \text{ and } f(x) \in H \Leftrightarrow f^{\leftarrow}_{|_\nu}(H) \in \tau \text{ and } x \in f^{\leftarrow}_{|_\nu}(H) \\ &\Leftrightarrow f^{\leftarrow}_{|_\nu}(H) \in \eta_{(X,\tau)}(x) \Leftrightarrow H \in \left(f^{\leftarrow}_{|_\nu}\right)^{\leftarrow}\left(\eta_{(X,\tau)}(x)\right) \\ &\Leftrightarrow H \in \left[Pt_{AG}\Omega_{AG}(f)\right]\left(\eta_{(X,\tau)}(x)\right). \end{aligned}$$

**Lemma 3.** *For every* ***Agop****-object* $(L, \leq, A)$*, the map* $e_{(L,\leq,A)} : L \to \Psi(L)$*, defined by* $e_{(L,\leq,A)}(a) = \Psi_a$*, is an* ***Agop****-morphism* $(L, \leq, A) \to \Omega_{AG}Pt_{AG}(L, \leq, A)$*. Moreover,* $\varepsilon = \left(e^{op}_{(L,\leq,A)}\right)_{(L,\leq,A)\in Ob(\mathbf{Agop})} : \Omega_{AG}Pt_{AG} \to id_{\mathbf{Agop}^{op}}$ *is a natural transformation.*

*Proof.* Consider first that $\Omega_{AG}Pt_{AG}(L, \leq, A) = \left(\Psi(L), \subseteq, \mathfrak{F}(L)(\cap)_{\Psi(L)}\right)$. In order to prove that $e_{(L,\leq,A)} : (L, \leq, A) \to \Omega_{AG}Pt_{AG}(L, \leq, A)$ is an **Agop**-morphism, we proceed as follows. For all $a, b \in L$ with $a \leq b$, since $\Psi_a \subseteq \Psi_b$ (by (F1) in Definition 4), $e_{(L,\leq,A)}$ is order-preserving. We also obtain from (F2) and (F3) in Definition 4 that $e_{(L,\leq,A)}$ preserves $\bot$ and $\top$. Furthermore, by making use of (F4) in Definition 4, we see that $\Psi_{a_1} \cap ... \cap \Psi_{a_n} = \Psi_{A(a_1,...,a_n)}$ for all $a_1, ..., a_n \in L$, and so

$$\begin{aligned} e_{(L,\leq,A)}(A(a_1, ..., a_n)) &= \Psi_{A(a_1,...,a_n)} = \Psi_{a_1} \cap ... \cap \Psi_{a_n} \\ &= e_{(L,\leq,A)}(a_1) \cap ... \cap e_{(L,\leq,A)}(a_n) \\ &= \mathfrak{F}(L)(\cap)_{\Psi(L)}\left(e_{(L,\leq,A)}(a_1), ..., e_{(L,\leq,A)}(a_n)\right). \end{aligned}$$

This completes the proof of the first part of the assertion. For the second part, the only property of $\varepsilon$ that we have to verify is its naturality, i.e. the commutativity of the diagram

$$\begin{array}{ccc} \Omega_{AG}Pt_{AG}(L, \leq, A) & \overset{\varepsilon_{(L,\leq,A)}}{\longrightarrow} & (L, \leq, A) \\ \Omega_{AG}Pt_{AG}(u) \downarrow & & \downarrow u \\ \Omega_{AG}Pt_{AG}(M, \leq, B) & \overset{\varepsilon_{(M,\leq,B)}}{\longrightarrow} & (M, \leq, B) \end{array} \tag{1}$$

for each $\mathbf{Agop}^{op}$-morphism $(L, \leq, A) \overset{u}{\to} (M, \leq, B)$. Since all arrows and all compositions in (1) are taken in $\mathbf{Agop}^{op}$, it can be simplified to be a rectangle

$$\begin{array}{ccc} (M, \leq, B) & \overset{e_{(M,\leq,B)}}{\longrightarrow} & \left(\Psi(M), \subseteq, \mathfrak{F}(M)(\cap)_{\Psi(M)}\right) \\ u^{op} \downarrow & & \downarrow Pt_{AG}(u)^{\leftarrow}_{|_{\Psi(M)}} \\ (L, \leq, A) & \overset{e_{(L,\leq,A)}}{\longrightarrow} & \left(\Psi(L), \subseteq, \mathfrak{F}(L)(\cap)_{\Psi(L)}\right), \end{array} \tag{2}$$

where all arrows and all compositions are performed in **Agop**. The commutativity of (2) is obtained as

$$\begin{aligned} \left[Pt_{AG}(u)^{\leftarrow}_{|_{\Psi(M)}} \circ e_{(M,\leq,B)}\right](b) &= [Pt_{AG}(u)]^{\leftarrow}(\Psi_b) = \Psi_{u^{op}(b)} \\ &= \left[e_{(L,\leq,A)} \circ u^{op}\right](b) \end{aligned}$$

for all $b \in M$.

**Theorem 1.** $(\eta, \varepsilon) : \Omega_{AG} \dashv Pt_{AG} : \boldsymbol{Agop}^{op} \to \boldsymbol{AGS}$ *is an adjoint situation.*

*Proof.* It is not difficult to check the adjunction identities

$$Pt_{AG}\left(\varepsilon_{(L,\leq,A)}\right) \circ \eta_{Pt_{AG}(L,\leq,A)} = id_{Pt_{AG}(L,\leq,A)},$$
$$\varepsilon_{\Omega_{AG}(X,\tau)} \circ \Omega_{AG}\left(\eta_{(X,\tau)}\right) = id_{\Omega_{AG}(X,\tau)}$$

for every **Agop**-object $(L, \leq, A)$ and every **AGS**-object $(X, \tau)$. Then, the required result follows immediately from Lemma 2 and Lemma 3.

*Remark 1.* Since $\Omega_{AG}(X,\tau) = (\tau, \subseteq, X(\cap)_\tau)$ is an object of **Smasagop** for every **AGS**-object $(X, \tau)$, the adjoint situation in Theorem 1 can be restricted to an adjoint situation $(\eta, \varepsilon^r) : \Omega^r_{AG} \dashv Pt^r_{AG} : \mathbf{Smasagop}^{op} \to \mathbf{AGS}$, where $\Omega^r_{AG}$ $(Pt^r_{AG})$ is the co-domain (the domain) restriction of $\Omega_{AG}$ $(Pt_{AG})$ and $\varepsilon^r_{(L,\leq,A)} = \varepsilon_{(L,\leq,A)}$ for every **Smasagop**-object $(L, \leq, A)$. An analogous adjoint situation can also be written for the category **Asagop** instead of **Smasagop**.

## 4 Spatial Aggregation Operators, Sober Aggregative Spaces and Their Duality

Spatiality and sobriety are two important notions that enable us to restrict the adjunction $\Omega_{AG} \dashv Pt_{AG}$ to an equivalence. To clarify this fact, we first start with their definitions:

**Definition 5.** *(i) An* ***Agop****-object* $(L, \leq, A)$ *is called spatial iff for all* $a, b \in L$ *with* $a \not\leq b$*, there exists a* $G \in \mathfrak{F}(L)$ *such that* $a \in G$ *and* $b \notin G$.

*(ii) An* ***AGS****-object* $(X, \tau)$ *is called sober iff for all* $\mathcal{U} \in \mathfrak{F}(\tau)$*, there exists a unique* $x \in X$ *such that* $\mathcal{U} = \tau(x)$.

**Proposition 4.** *Let* $(L, \leq, A)$ *be an* ***Agop****-object, and* $(X, \tau)$ *an* ***AGS****-object.*

*(i)* $(L, \leq, A)$ *is spatial iff* $e_{(L,\leq,A)} : (L, \leq, A) \to \Omega_{AG} Pt_{AG}(L, \leq, A)$ *is an* ***Agop****-isomorphism.*

*(ii)* $(L, \leq, A)$ *is spatial iff* $(L, \leq, A)$ *is isomorphic to* $(\nu, \subseteq, Y(\cap)_\nu)$ *for some aggregative space* $(Y, \nu)$.

*(iii)* $(X, \tau)$ *is sober iff* $\eta_{(X,\tau)} : (X, \tau) \to Pt_{AG}\Omega_{AG}(X, \tau)$ *is an* ***AGS****-isomorphism.*

*Proof.* (i) Note first that $(L, \leq, A)$ is spatial iff for all $a, b \in L$, $\Psi_a \subseteq \Psi_b$ implies $a \leq b$. Now, by assuming spatiality of $(L, \leq, A)$, this equivalence directly gives the injectivity of the underlying set map of $e_{(L,\leq,A)}$, and so does its bijectivity. It is easy to check that $e^{-1}_{(L,\leq,A)} : \Omega_{AG} Pt_{AG}(L, \leq, A) \to (L, \leq, A)$ is an **Agop**-morphism, and so $e_{(L,\leq,A)} : (L, \leq, A) \to \Omega_{AG} Pt_{AG}(L, \leq, A)$ is an **Agop**-isomorphism. Conversely, if $e_{(L,\leq,A)} : (L, \leq, A) \to \Omega_{AG} Pt_{AG}(L, \leq, A)$ is an **Agop**-isomorphism, then since $e^{-1}_{(L,\leq,A)} : \Omega_{AG} Pt_{AG}(L, \leq, A) \to (L, \leq, A)$ is an **Agop**-morphism, $\Psi_a \subseteq \Psi_b$ implies $a = e^{-1}_{(L,\leq,A)}(\Psi_a) \leq e^{-1}_{(L,\leq,A)}(\Psi_b) = b$ for all $a, b \in L$, so $(L, \leq, A)$ is spatial.

(ii) If $(L, \leq, A)$ is spatial, then we have from (i) that $(L, \leq, A)$ is isomorphic to $(\nu, \subseteq, Y(\cap)_\nu)$ for the aggreative space $(Y, \nu) = Pt_{AG}(L, \leq, A)$. Conversely, suppose $(L, \leq, A)$ is isomorphic to $(\nu, \subseteq, Y(\cap)_\nu)$ for some aggregative space $(Y, \nu)$, i.e. there exists an **Agop**-isomorphism $u : (L, \leq, A) \to (\nu, \subseteq, Y(\cap)_\nu)$. Then, for $a, b \in L$ with $a \not\leq b$, since $u(a) \not\subseteq u(b)$, there exists at least one $z \in Y$ such that $z \in u(a)$ and $z \notin u(b)$. It is clear that $u(a) \in \nu(z)$ and $u(b) \notin \nu(z)$, and so $a \in u^{\leftarrow}(\nu(z))$ and $b \notin u^{\leftarrow}(\nu(z))$. Hence, we obtain the spatiality of $(L, \leq, A)$ from the fact that $u^{\leftarrow}(\nu(z))$ is a filter of $(L, \leq, A)$.

(iii) follows from that for a given **AGS**-object $(X, \tau)$, $(X, \tau)$ is sober iff the underlying set map of $\eta_{(X,\tau)}$ is a bijection iff $\eta_{(X,\tau)} : (X, \tau) \to Pt_{AG}\Omega_{AG}(X, \tau)$ is an **AGS**-isomorphism.

**Corollary 1.** *The full subcategory* ***SpatAgop*** *of* ***Agop*** *of all spatial objects is dually equivalent to the full subcategory* ***SobAGS*** *of* ***AGS*** *of all sober objects.*

*Proof.* Since Proposition 4 (i) and (iii) verify that $Fix(\varepsilon) =$ **SpatAgop** and $Fix(\eta) =$ **SobAGS**, the assertion follows from Theorem 1 and Proposition 1 (i).

**Proposition 5.** ***SpatAgop*** *and* ***SobAGS*** *are reflective subcategories of* ***Agop*** *and of* ***AGS*** *with reflectors* $\Omega^{op}_{AG} Pt^{op}_{AG}$ *and* $Pt_{AG}\Omega_{AG}$*, and the reflection arrows* $e_{(L,\leq,A)}$ *and* $\eta_{(X,\tau)}$*, respectively.*

*Proof.* Since $e_{(L,\leq,A)} = \varepsilon^{op}_{(L,\leq,A)}$ is obviously an epimorphism in **Agop**, and $Fix(\varepsilon) =$ **SpatAgop** and $Fix(\eta) =$ **SobAGS**, Proposition 1 (ii) directly yields the claimed result.

**Proposition 6.** *Let* $(L, \leq, A)$ *be an* ***Agop****-object, and* $(X, \tau)$ *an* ***AGS****-object.*
*(i)* $\Omega_{AG}(X, \tau)$ *is spatial, (ii)* $Pt_{AG}(L, \leq, A)$ *is sober.*

*Proof.* (i) is immediate from Proposition 4 (ii). To see (ii), let us first consider the fact that $Pt_{AG}(L, \leq, A) = (\mathfrak{F}(L), \Psi(L))$. Then, the sobriety of $Pt_{AG}(L, \leq, A)$ follows from the observation that for all $\mathcal{U} \in \mathfrak{F}(\Psi(L))$, $G = \{a \in L \mid \Psi_a \in \mathcal{U}\}$ is the unique element of $\mathfrak{F}(L)$ with the property that $\mathcal{U} = [\Psi(L)](G)$.

**Proposition 7.** *The full subcategory* ***SpatAsagop*** *of* ***Asagop*** *of all spatial objects, the full subcategory* ***SpatSmasagop*** *of* ***Smasagop*** *of all spatial objects and* ***SpatAgop*** *are equivalent to each others.*

*Proof.* Since **SpatSmasagop** is a full subcategory of **SpatAgop**, the inclusion functor **SpatSmasagop** $\hookrightarrow$ **SpatAgop** is a full and faithful functor. For every **SpatAgop**-object $(L, \leq, A)$, by Proposition 4 (i) and Proposition 6 (i), $\Omega_{AG}Pt_{AG}(L, \leq, A)$ is a **SpatSmasagop**-object, and $e_{(L,\leq,A)} : (L, \leq, A) \to \Omega_{AG}Pt_{AG}(L, \leq, A)$ is a **SpatAgop**-isomorphism. This proves that the inclusion functor **SpatSmasagop** $\hookrightarrow$ **SpatAgop** is isomorphism-dense, and hence an equivalence. Similarly, the inclusion functor **SpatAsagop** $\hookrightarrow$ **SpatAgop** is an equivalence, which completes the proof.

**Corollary 2.** $\boldsymbol{SpatAgop}^{op} \sim \boldsymbol{SpatAsagop}^{op} \sim \boldsymbol{SpatSmasagop}^{op} \sim \boldsymbol{SobAGS}$.

## References


1. Adámek, J., Herrlich, H., Strecker, G.E.: Abstract and Concrete Categories. Wiley, New York (1990)
2. Banaschewski, B., Bruns, G.: The Fundamental Duality of Partially Ordered Sets. Order 5, 61–74 (1988)
3. Calvo, T., Kolesárová, A., Komorníková, M., Mesiar, R.: Aggregation Operators: Properties, Classes and Construction Methods. In: Calvo, T., et al. (eds.) Aggregation Operators. New Trends and Applications, vol. 97, pp. 3–104. Physica-Verlag, Heidelberg (2002)
4. De Baets, B., Mesiar, R.: Triangular Norms on Product Lattices. Fuzzy Sets and Systems 104, 61–75 (1999)
5. Demirci, M.: Aggregation Operators on Partially Ordered Sets and Their Categorical Foundations. Kybernetika 42, 261–277 (2006)
6. Demirci, M.: $(\mathcal{Z}_1, \mathcal{Z}_2)$ -complete Partially Ordered Sets and Their Representations by $\mathcal{Q}$ -spaces. Appl. Categ. Struc. 21, 703–723 (2013)
7. Demirci, M.: Fundamental Duality of Abstract Categories and Its Applications. Fuzzy Sets and Systems, `http://dx.doi.org/10.1016/j.fss.2013.08.015`
8. Deschrijver, G., Kerre, G., Implicators Based, E.E.: on Binary Aggregation Operators in Interval-valued Fuzzy Set Theory. Fuzzy Sets and Systems 153, 229–248 (2005)
9. Johnstone, P.T.: Stone Spaces. Cambridge University Press, Cambridge (1986)
10. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Trends in Logic, vol. 8. Kluwer Academic Publishers, Dordrecht (2000)
11. Komorníková, M., Mesiar, R.: Aggregation Functions on Bounded Partially Ordered Sets and Their Classification. Fuzzy Sets and Systems 175, 48–56 (2011)
12. Lázaro, J., Calvo, T.: XAO Operators - The Interval Universe. In: Proc. 4th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2005), pp. 198–203 (2005)
13. Mesiar, R., Komorníková, M.: Classification of Aggregation Functions on Bounded Partially Ordered Sets. In: Proceedings of the SISY 2010, Subotica, September 10-11, pp. 13–16 (2010)
14. Saminger, S., Mesiar, R., Bodenhofer, U.: Domination of Aggregation Operators and Preservation of Transitivity. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 10, 11–35 (2002)
15. Saminger, S.: On ordinal sums of triangular norms on bounded lattices. Fuzzy Sets and Systems 157, 1403–1416 (2006)
16. Saminger-Platz, S., Klement, E.P., Mesiar, R.: On extensions of triangular norms on bounded lattices. Indagationes Mathematicae 19, 135–150 (2008)
17. Wanga, Z., Fanga, J.: Residual operations of left and right uninorms on a complete lattice. Fuzzy Sets and Systems 160, 22–31 (2009)
18. Sua, Y., Wanga, Z.: Pseudo-uninorms and coimplications on a complete lattice. Fuzzy Sets and Systems 224, 53–62 (2013)

# Smart Fuzzy Weighted Averages of Information Elicited through Fuzzy Numbers

Andrea Capotorti and Gianna Figá-Talamanca

Universitá degli Studi di Perugia, Italy
capot@dmi.unipg.it, gianna.figatalamanca@unipg.it

**Abstract.** We illustrate a preliminary proposal of weighted fuzzy averages between two membership functions. Conflicts, as well as agreements, between the different sources of information in the two new operators are endogenously embedded inside the average weights. The proposal is motivated by the practical problem of assessing the fuzzy volatility parameter in the Black and Scholes environment via alternative estimators.

**Keywords:** merging, aggregation, fuzzy mean.

## 1 Introduction and Motivation

In [2] we introduced a methodology for membership elicitation on parameters. The goal was to estimate the hidden volatility parameter $\sigma$ of a risky asset through both the historical volatility estimator $\widehat{\sigma}$, based on a sample of log-returns of the asset itself, and the estimator $\nu = \text{VIX}/100$, based on VIX which is a volatility index obtained through a set of prices for options written on the asset. Thanks to the interpretation of membership functions as coherent conditional probability assessments (see [3,4]) integrated with observational data and expert evaluations, we were able in some cases to elicit proper membership functions for the volatility parameter based on each single estimator, while in another case two memberships were considered possible. Moreover, results were obtained through probability-possibility transformation of specific simulation distributions. Thus, the peculiarity of the proposal was to deal with implicit and alternative sources of information, while one of the open problem was to find proper fusion operators.

In literature it is known that the choice of a fusion operator, given the variety of information items, is not unique and heavily context-dependent. Classes of aggregation functions covered include triangular norms and conorms, copulas, means and averages, and those based on nonadditive integrals [11]. A main characteristic of the aggregation functions is that they are used in a large number of areas and disciplines, leading to a strong demand for a wide variety of aggregation functions with predictable and tailored properties [12], [13]. Authors in [15] affirm that there are more than 90 different fuzzy operators proposed in the literature for fuzzy set operations. The role of fuzzy sets in merging information can be understood in two ways: either as a tool for extending estimation techniques to fuzzy data (this is done applying the extension principle to classical

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 466–475, 2014.

estimators, and methods of fuzzy arithmetics - see [5] for a survey); or as a tool for combining possibility distributions that represent imprecise pieces of information (then fuzzy set-theoretic operations are instrumental for this purpose - see [9] for a survey).

In view of this dichotomy, the role of standard aggregation operations like arithmetic mean is twofold. It is a basic operation for estimation and also a fuzzy set-theoretic connective. A bridge between the "estimation" and the "fusion" views of merging information is ensured using the concept of *constrained merging* [8, §6.6.2]. We borrow from it the motivation of including a "smart" component in the averaging process to address conflicts in the data to be fused, but, contrarily to the original "intelligent" proposal, without the introduction of an exogenous "combinability function". We need two different kinds of fusion operators: one for merging conjointly the values stemming from the different estimators; and another that disjointly considers different possibilities or distribution models. Our operators are weighted averages where conflicts, as well as agreements, are endogenously embedded on the average weights; for the reasons mentioned above the choices for the weights we suggest here are deeply motivated by the practical problem at hand. The difference between the two proposals is in which direction there is a deformation of the arithmetic mean: for merging of joint information distortion is toward canonical conjunction, i.e. min, while for merging of alternative information distortion is toward canonical disjunction, i.e. max.

The rest of the paper is organized as follows: next section briefly refreshes main fuzzy membership notions and introduces basic notations for our purposes, while subsequent Section 3 defines our weighted averages proposals. Section 4 illustrates the numerical applications to the original practical problem of elicitation of a single membership function for the fuzzy volatility parameter $\widetilde{\sigma}$ and its consequences on the option pricing. A similarity comparison with crisp bid-ask prices is also performed. Section 5 briefly concludes the contribution.

## 2 Notation

Given our goal of parameter estimation, for the sequel we will consider real valued quantities. We recall that a membership function $\mu : \mathbb{R} \to [0, 1]$ of the fuzzy set of possible values of a random variable $X$ can be viewed as a possibility distribution (see e.g. [17]). In particular, the subset $\mu^S = \{x \in \mathbb{R} : \mu(x) > 0\}$ is named the "support" of the membership while the subset $\mu^1 = \{x \in \mathbb{R} : \mu(x) = 1\}$ is its "core". Membership functions are fully characterized by their (dual) representation through $\alpha$-cuts $\mu^\alpha = \{x \in \mathbb{R} : \mu(x) \geq \alpha\}, \quad \alpha \in [0, 1]$. The $\alpha$ value can be conveniently interpreted as 1 minus the lower bound of the probability that quantity $X$ hits $\mu^\alpha$. Then the possibility distribution is viewed as the family of probability measures ([6]): $\mathcal{P} = \{\text{prob. distr. } P : P(X \in \mu^\alpha) \geq 1 - \alpha\}$. In [2] we were able to elicit membership functions through probability-possibility transformations ([7]) induced by confidence intervals around the median of specific simulating distributions; we got so called "fuzzy numbers", i.e. unimodal

membership functions with nested $\alpha$-cuts. Hence, each $\mu$ we consider has an increasing left branch $\mu_l$ and a decreasing right one $\mu_r$ and each $\alpha$-cut is identified by an interval $[\mu_l^\alpha, \mu_r^\alpha]$ in the extended reals $\widetilde{\mathbb{R}}$.

Aggregations are performed between $\alpha$-cuts, so we always deal with two intervals, possibly degenerate, $[\mu 1_l^\alpha, \mu 1_r^\alpha]$ and $[\mu 2_l^\alpha, \mu 2_r^\alpha]$. From these it is immediate to define their four characteristic values

$$\mu_{lO}^\alpha = \min\{\mu 1_l^\alpha, \mu 2_l^\alpha\} \quad \mu_{lI}^\alpha = \max\{\mu 1_l^\alpha, \mu 2_l^\alpha\}; \tag{1}$$

$$\mu_{rI}^\alpha = \min\{\mu 1_r^\alpha, \mu 2_r^\alpha\} \quad \mu_{rO}^\alpha = \max\{\mu 1_r^\alpha, \mu 2_r^\alpha\}, \tag{2}$$

and their lengths

$$\Delta_1^\alpha = \mu 1_r^\alpha - \mu 1_l^\alpha \qquad \Delta_2^\alpha = \mu 2_r^\alpha - \mu 2_l^\alpha \qquad \Delta^\alpha = \mu_{rO}^\alpha - \mu_{lO}^\alpha; \tag{3}$$

where the subscript O refers to the "outer" values, while the subscript I to the "inner" ones (see e.g. Fig.1).

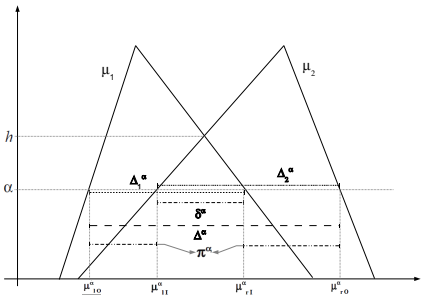



**Fig. 1.** Characteristic values for the merging of two $\alpha$-cuts

A crucial value for our proposal is the "height" $h$ of the intersection between $\mu 1$ and $\mu 2$, i.e.

$$h = \max\{\alpha : [\mu 1_l^\alpha, \mu 1_r^\alpha] \cap [\mu 2_l^\alpha, \mu 2_r^\alpha] \neq \emptyset\} \tag{4}$$

whenever the two memberships overlaps, while $h = 0$ if $\mu 1$ and $\mu 2$ are incompatible. Other involved quantities are

$$\delta^\alpha = |\mu_{rI}^\alpha - \mu_{lI}^\alpha| \tag{5}$$

that measures the width of the intersection if the two $\alpha$-cuts overlaps or, otherwise, the minimal distance between them; and

$$\pi^\alpha = \Delta^\alpha - \delta^\alpha \tag{6}$$

that measure the length of the parts outside the (possible) intersection.

With such quantities, for the levels $\alpha \leq h$ we can define the relative contributions $\epsilon_l^\alpha$ and $\epsilon_r^\alpha$ of the inner memberships $\mu_{lI}, \mu_{rI}$ to the intersection as:

$$\epsilon_l^\alpha = \frac{\frac{\delta^\alpha}{\Delta_{lI}^\alpha}}{\frac{\delta^\alpha}{\Delta_{lI}^\alpha} + \frac{\delta^\alpha}{\Delta_{lO}^\alpha}} = \frac{\Delta_{lO}^\alpha}{\Delta_1^\alpha + \Delta_2^\alpha}; \tag{7}$$

$$\epsilon_r^\alpha = \frac{\frac{\delta^\alpha}{\Delta_{rI}^\alpha}}{\frac{\delta^\alpha}{\Delta_{rI}^\alpha} + \frac{\delta^\alpha}{\Delta_{rO}^\alpha}} = \frac{\Delta_{rO}^\alpha}{\Delta_1^\alpha + \Delta_2^\alpha}; \tag{8}$$

with

$$\Delta_{lI}^\alpha = \begin{cases} \Delta_1^\alpha & \text{if } \mu_{lI}^\alpha = \mu 1_l^\alpha \\ \Delta_2^\alpha & \text{if } \mu_{lI}^\alpha = \mu 2_l^\alpha \end{cases} \quad \Delta_{lO}^\alpha = \begin{cases} \Delta_1^\alpha & \text{if } \mu_{lO}^\alpha = \mu 1_l^\alpha \\ \Delta_2^\alpha & \text{if } \mu_{lO}^\alpha = \mu 2_l^\alpha \end{cases} \tag{9}$$

and, similarly,

$$\Delta_{rI}^\alpha = \begin{cases} \Delta_1^\alpha & \text{if } \mu_{rI}^\alpha = \mu 1_r^\alpha \\ \Delta_2^\alpha & \text{if } \mu_{rI}^\alpha = \mu 2_r^\alpha \end{cases} \quad \Delta_{rO}^\alpha = \begin{cases} \Delta_1^\alpha & \text{if } \mu_{rO}^\alpha = \mu 1_r^\alpha \\ \Delta_2^\alpha & \text{if } \mu_{rO}^\alpha = \mu 2_r^\alpha \end{cases}. \tag{10}$$

## 3 A Proposal of Two Smart Weighted Averages

We propose two new binary operations $\overline{\wedge}$ and $\underline{\vee}$ to average in a conjunctive or in a disjunctive way, respectively, different information $\mu 1$ and $\mu 2$. Both generalize, by deformation, the usual arithmetic mean between two fuzzy numbers: $\overline{\wedge}$ deforms the arithmetic mean toward the min conjunction operator, while $\underline{\vee}$ toward the max disjunction operator.

Hence we define our generalized conjunction level-wise by setting as $\alpha$-cut $(\mu 1 \,\overline{\wedge}\, \mu 2)^\alpha$ the interval

$$[(\mu 1 \,\overline{\wedge}\, \mu 2)_l^\alpha, (\mu 1 \,\overline{\wedge}\, \mu 2)_r^\alpha] = [wl^\alpha \mu_{lI}^\alpha + (1 - wl^\alpha)\mu_{lO}^\alpha \,,\, wr^\alpha \mu_{rI}^\alpha + (1 - wr^\alpha)\mu_{rO}^\alpha] \tag{11}$$

with weights:
for $\alpha \leq h$

$$wl^\alpha = \frac{1}{2} + \frac{\epsilon_l^\alpha}{2} \quad , \quad wr^\alpha = \frac{1}{2} + \frac{\epsilon_r^\alpha}{2}, \tag{12}$$

for $\alpha > h$,

$$wl^\alpha = \frac{(\mu 1 \,\overline{\wedge}\, \mu 2)_l^h - \mu_{lO}^\alpha + k(M_l^\alpha - M_l^h) + \theta_l(\alpha)}{(\mu_{lI}^\alpha - \mu_{lO}^\alpha)} \tag{13}$$

$$wr^\alpha = \frac{(\mu 1 \,\overline{\wedge}\, \mu 2)_r^h - \mu_{rI}^\alpha + k(M_r^\alpha - M_r^h) - \theta_r(\alpha)}{(\mu_{rO}^\alpha - \mu_{rI}^\alpha)} \tag{14}$$

with $[M_l^\alpha, M_r^\alpha]$ the $\alpha$-cut of the arithmetic fuzzy mean; $\theta_l$ and $\theta_r$ specific quadratic functions used to emphasize the deformation, and $k = \frac{(\mu 1 \overline{\wedge} \mu 2)_r^h - (\mu 1 \overline{\wedge} \mu 2)_l^h}{M_r^h - M_l^h}$ a scale factor. It is important to remark that, since the intersections between the $\alpha$-cuts are empty for $\alpha > h$, the choice of the weights in that case just resumes our operator to the arithmetic mean, but shifted and deformed to be

"glued" with the lower levels and to emphasize the contradiction between the two sources.

Similarly, our generalized disjunction $\alpha$-cuts $(\mu 1 \veebar \mu 2)^\alpha$ is defined as

$$[W^\alpha \mu^\alpha_{lO} + (1 - W^\alpha)\mu^\alpha_{lI}\,,\, W^\alpha \mu^\alpha_{rO} + (1 - W^\alpha)\mu^\alpha_{rI}] \tag{15}$$

with weights

$$W^\alpha = \frac{1 + \frac{(1-\alpha)\,\pi^\alpha}{\Delta^\alpha}}{2}\,. \tag{16}$$

We have already underlined that the main goal of the averaging operators $\barwedge$ and $\veebar$ is to deform usual fuzzy arithmetic mean toward min and max connectives, respectively. This realizes by the inter-change in equations (11) and (15) among the extremes. In fact in (11) weights $wl^\alpha$ and $wr^\alpha$ deform the results toward the "inner" part, through (12), until there is an overlap between the $\alpha$-cuts, and abroad from it, through (13,14), otherwise. On the contrary in (15) weights $W^\alpha$ always deform the average towards the "outer" part, as much as there is "contradiction" between the two memberships. Other properties of $\barwedge$ and $\veebar$ are the closure (both averages of two fuzzy numbers produce a fuzzy number), the idempotence and the symmetry. It is easy to find examples of non-associativity of $\barwedge$ and $\veebar$, but virtually no averaging operation is associative because it is known [10] that the only associative averaging operations are of the form median.

## 4 Applied Example

As already stressed in the Introduction, the proposed weighted averages $\barwedge$ and $\veebar$ were motivated by the need left unresolved in [2] of an implicit assessment of fuzzy volatility in the Black and Scholes environment based on two different estimators $\widehat{\sigma}$ and $\nu$, and on different simulating models for searched parameter $\sigma$. In particular, for each estimator, different scenarios are considered on the base of historical data and experts evaluations. For each scenario it was possible to build a pseudo-membership for the considered estimator by coherent extension of a-priori information and likelihood values stemming from specific simulation distribution of the unknown parameter. At the end, observed values of the estimators permitted to select most plausible scenarios, that could be a single one if there were sure dominance of one scenario over the others, or more than one if dominance was partial. For each scenario a probability-possibility transformation of the associated simulating distributions gave as results different membership functions. The adopted simulating distributions for $\sigma$ were the uniform, the log-normal and the gamma densities, with parameters determined by the different scenarios characteristics. Hence, we have to face several merging requirements: among memberships associated to different selected scenarios, among memberships stemming from different simulating functions and between memberships associated to the two different estimators $\widehat{\sigma}$ and $\nu$.

### 4.1 Elicitation of Fuzzy Volatility

As preliminary illustrative results let us show a prototypical situation (corresponding to "Case 2" in [2]) where for at least one estimator there is more than one plausible scenario and different simulating models produce quite different outputs, though at the end the two sources give quite agreeing results. In particular, for such "Case 2", associated to $\widehat{\sigma}_{obs} = 0.16$ the Log-Normal simulating model furnished two alternative scenarios (the "medium" or the "high") among the five considered, while the other two models agreed in selecting only the "medium" one. Here, by transforming the three simulating distributions we obtain the memberships reported in Fig.2 (a) where for the Log-Normal the two alternative memberships has been already merged through (15). About the

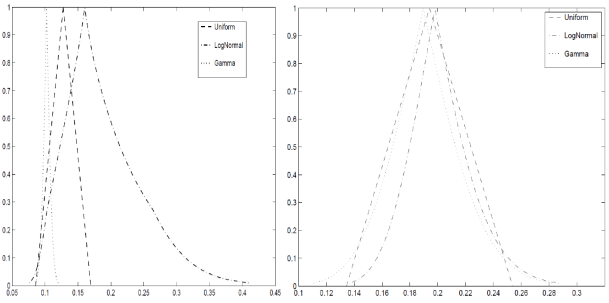



**Fig. 2.** Membership functions for Case 2 representing scenarios stemming from different simulating distributions, as selected by $\widehat{\sigma}_{obs} = 0.16$ (a) or by $\nu_{obs} = 0.19$ (b)

other estimator $\nu$, its observed value $\nu_{obs} = 0.19$ always led to the selection of the "medium" scenario, obtaining the three memberships plotted in Fig.2 (b).

Since the simulating models are alternative, for both groups we can apply level-wise the weighted average (15) just between the two most contradictory memberships, since the third remains fully covered by the others. At this point we have two fuzzy numbers representative of the two sources $\mu_{\widehat{\sigma}_{obs}}$ and $\mu_{\nu_{obs}}$ which can be merged in a conjunctive way obtaining the final result $\mu_\sigma = \mu_{\widehat{\sigma}_{obs}} \,\overline{\wedge}\, \mu_{\nu_{obs}}$ reported in Fig.3.

### 4.2 From Elicitation of the Fuzzy Volatility to Fuzzy Option Pricing

It is well known that under the assumptions in Black and Scholes ([1]), a closed formula is available for the price of European Call Options, given by

$$C(t, S, r, \sigma, K, T) = SN(d_1) - e^{-r(T-t)}KN(d_2), \tag{17}$$

with

$$d_1 = \frac{\log(S/K) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T-t}, \tag{18}$$

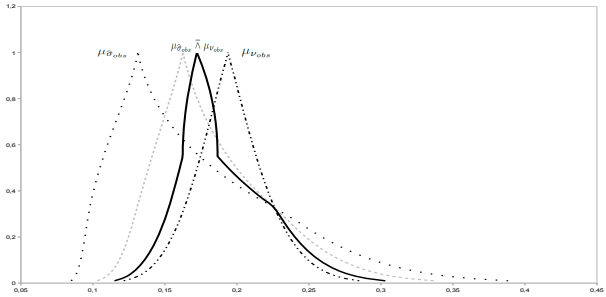


**Fig. 3.** Final membership function (solid) for $\sigma$ in Case 2 obtained as $\mu_\sigma = \mu_{\widehat{\sigma}_{obs}} \overline{\wedge} \mu_{\nu_{obs}}$ of the information coming from $\widehat{\sigma}$ (dotted) and $\nu$ (dashed-dotted). Arithmetic mean (grey dashed) has been also reported for comparison.

where $K, T$ are the strike price and the maturity of the Option, respectively, S is the price in $t$ of the underlying asset and $r, \sigma$ are model parameters representing the constant risk free continuously compounded rate and the volatility of the asset, and $N(\cdot)$ is the standard Normal cumulative function. Let us consider function $C$ as a function of the volatility parameter only, assuming the other inputs as constant values i.e. $c = C(t, S, r, \sigma, K, T) = C(\sigma)$. Assuming that parameter $\sigma$ is modeled as a a fuzzy number $\widetilde{\sigma}$, it is possible to detect the propagation of uncertainty from the volatility parameter to the option price by defining the fuzzy extension $\widetilde{c} = C(\widetilde{\sigma})$ of $c = C(\sigma)$; if the volatility $\sigma$ is a fuzzy number $\widetilde{\sigma}$ described through its $\alpha$-cuts $[\widetilde{\sigma}_l^\alpha, \widetilde{\sigma}_r^\alpha]$, for each level $\alpha$, then the option price $c$ is still a fuzzy number $\widetilde{c}$, also described by its $\alpha$-cuts $[\widetilde{c}_l^\alpha, \widetilde{c}_r^\alpha]$. To obtain the fuzzy extension of $C$ to normal upper semi-continuous fuzzy intervals one may apply the methodology as in [14], based on the solution of the box-constrained optimization problems

$$\begin{cases} \widetilde{c}_l^\alpha = \min\{C(\sigma) | \sigma \in [\widetilde{\sigma}_l^\alpha, \widetilde{\sigma}_r^\alpha]\} \\ \widetilde{c}_r^\alpha = \max\{C(\sigma) | \sigma \in [\widetilde{\sigma}_l^\alpha, \widetilde{\sigma}_r^\alpha]\}. \end{cases} \tag{19}$$

Since $C$ is a strictly increasing function in $\sigma$ we easily obtain

$$\begin{cases} \widetilde{c}_l^\alpha = C(\widetilde{\sigma}_l^\alpha) \\ \widetilde{c}_r^\alpha = C(\widetilde{\sigma}_r^\alpha). \end{cases} \tag{20}$$

### 4.3 Empirical Application

According to the fuzzy number obtained by suitably merging information on volatility as in Subsection 4.1, we compute the corresponding fuzzy option prices for SPX options written on the S&P500 Index. We considered options traded on October 21st, 2010: a maximum of 50 strike prices were available (for the one-month to maturity options) as well as 11 different expiration dates for a total of 168 options. The underlying price on October, 21st 2010 was S=1180.26. In order to asses the empirical significance of fuzzy option prices computed via our

approach we need a proper comparison with the market bid and ask prices for the corresponding options. Besides, a selection criteria is needed to identify a set of more representative options on which to base our empirical exercise; it is well known that the more an option is traded the more its price may be interpreted as as an equilibrium price between supply and demand. For this reason, we compute, for each expiration date available, the mean trading volume obtained as the ratio of the total trading volume on options with that maturity and the total number of options with that maturity. We select, this way, 37 options.

To the end of comparing fuzzy option prices to market prices, we compute a suitably defined measure of fuzzy distances between the Black and Scholes fuzzy prices and the market bid-ask prices thought as crisp intervals, where the prices can be located with a step membership function with value 1 in the bid-ask interval and value 0 otherwise. For such purpose we consider two different fuzzy distance measures:

**(a)** the well known modified Bhattacharyya distance (see e.g. [18]) and
**(b)** the usual fuzzy similarity defined through min as $t$-norm and max as $t$-conorm (obtained also in [16] as special case of general similarities based on coherent conditional probabilities).

Further, to evaluate the added value of our merging approach with respect to usual fuzzy merging, we compute the distance/similarity measures also for the fuzzy option prices obtained applying the extension principle in the case when the fuzzy volatility parameter is modeled as the fuzzy arithmetic mean between memberships of $\widehat{\sigma}$ and $\nu$ (see a comparison of the two pricing techniques in Fig. 4). In 24 cases out of 37 the distance (a) is smaller for the fuzzy option prices obtained by our proper merging rather than by fuzzy arithmetic mean (see Fig. 5 (a)). Consistent results are obtained through the computation of the fuzzy similarity (b) which is larger for our fuzzy merging in 23 out of 24 cases above (see Fig. 5 (b)).

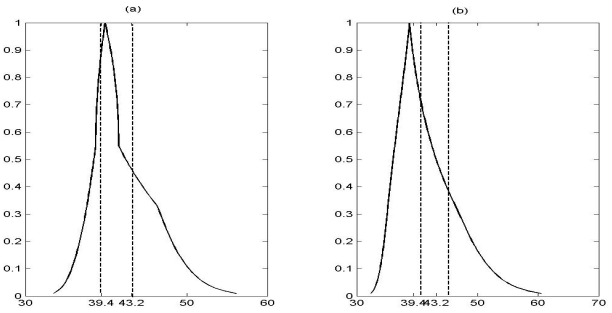



**Fig. 4.** Memberships of the SPX option price consequent of fuzzy volatility obtained by our merging (a) or by fuzzy arithmetic mean (b) of $\widehat{\sigma}$ and $\nu$; both for T=30 days and K=1150 and compared to the Bid-Ask crisp interval

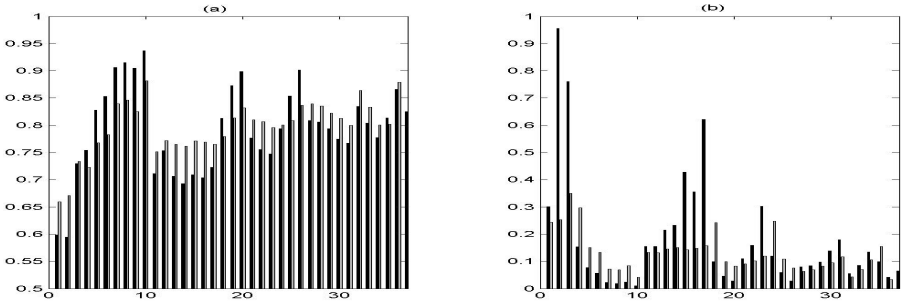



**Fig. 5.** Bhattacharyya distances (a) and fuzzy similarities (b) between the Bid_Ask crisp intervals and Fuzzy Option Prices obtained by our merging (black) or through the fuzzy arithmetic mean (grey)

## 5 Conclusion

We have illustrated a preliminary study of two weighted averages between membership functions that try to encompass in the usual fuzzy arithmetic mean the agreement or the contradiction of two heterogeneous sources of information. Formal properties of the two proposed operators $\overline{\wedge}$ and $\underline{\vee}$ must be fully investigated and practical consequences fully analyzed. Anyhow the first empirical results we have shown here seem to be promising, in particular with respect to the applicability in very different situations and the capability of conciliating quite different information.

**Acknowledgement.** This work was partially supported by Italian Ministry of Education, University and Research funding of Research Projects of National Interest (PRIN 2010-11) under grant `2010FP79LR_003` *"Logical methods of information management"*.

## References

1. Black, F., Scholes, M.: The pricing of options and corporate liabilities. Journal of Political Economy 81(3), 637–654 (1973)
2. Capotorti, A., Figá-Talamanca, G.: On an implicit assessment of fuzzy volatility in the Black and Scholes environment. Fuzzy Sets and Systems 223, 59–71 (2013)
3. Coletti, G., Scozzafava, R.: Conditional probability, fuzzy sets, and possibility: a unifying view. Fuzzy Sets and Systems 144, 227–249 (2004)
4. Coletti, G., Scozzafava, R.: Conditional probability and fuzzy information. Computational Statistics and Data Analysis 51(1), 115–132 (2006)
5. Dubois, D., Kerre, E., Mesiar, R.: Fuzzy interval analysis. In: Dubois, D., Prade, H. (eds.) Fundamentals of Fuzzy Sets. Handbook Series of Fuzzy Sets, pp. 483–582 (2000)

6. Dubois, D., Prade, H.: When Upper Probabilities Are Possibility Measures. Fuzzy Sets and Systems 49, 65–74 (1992)
7. Dubois, D., Prade, H., Sandri, S.: On Possibility/Probability Transformations. In: Lowen, R., Roubens, M. (eds.) Fuzzy Logic, pp. 103–112. Kluwer Academic Publishers, Dordrecht (1993)
8. Dubois, D., Prade, H., Yager, R.R.: Merging fuzzy information. In: Bezdek, J.C., Dubois, D., Prade, H. (eds.) Fuzzy Sets in Approximate Reasoning and Information Systems. The Handbooks of Fuzzy Sets Series, pp. 335–401. Kluwer Academic Publishers, Dordrecht (1999)
9. Fodor, J., Yager, R.: Fuzzy set-theoretic operators and quantifiers. In: Dubois, D., Prade, H. (eds.) Fundamentals of Fuzzy Sets. Handbook Series of Fuzzy Sets, pp. 125–193 (2000)
10. Fung, L.W., Fu, K.S.: An axiomatic approach to rational decision-making in a fuzzy environment. In: Zadeh, L.A., Fu, K.S., Tanaka, K., Shimura, M. (eds.) Fuzzy Sets and Their Application to Cognitive and Decision Processes, pp. 227–256. Academic Press, New York (1975)
11. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation Functions. Encyclopedia of Mathematics and its Applications, vol. 127. Cambridge University Press (2009)
12. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation functions: Means. Information Sciences 181, 1–22 (2011)
13. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation functions: construction methods, conjunctive, disjunctive and mixed classes. Information Sciences 181, 23–43 (2011)
14. Guerra, M.L., Sorini, L., Stefanini, L.: Parametrized Fuzzy Numbers for Option Pricing. In: 2007 IEEE Conference on Fuzzy Systems, vol. 1, pp. 728–733 (2007)
15. Li, H.-X., Yen, V.C.: Fuzzy Sets and Fuzzy Decision-making. CDC Press, Boca Raton (1995)
16. Scozzafava, R., Vantaggi, B.: Fuzzy Inclusion and Similarity Through Coherent Conditional Probability. Fuzzy Sets and Systems 160(3), 292–305 (2009)
17. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
18. Zwick, R., Carlstein, E., Budescu, D.V.: Mesures of Symilarity among Fuzzy Concepts: A Comparative Analysis. Int. J. of Approximate Reasoning 1, 221–242 (1987)

# Analytical Hierarchy Process under Group Decision Making with Some Induced Aggregation Operators

Binyamin Yusoff[1] and José M. Merigó-Lindahl[2,3]

[1] Department of Mathematical Economics and Finance, University of Barcelona,
Av. Diagonal 690, 08034, Barcelona, Spain
[2] Manchester Business School, University of Manchester,
Booth Street West, M15 6PB Manchester, UK
[3] Risk Center, University of Barcelona,
Av. Diagonal 690, 08034, Barcelona, Spain
binyaminy@yahoo.com, jmerigo@ub.edu

**Abstract.** This paper focuses on the extension of Analytical Hierarchy Process under Group Decision Making (AHP-GDM) with some induced aggregation operators. This extension generalizes the aggregation process used in AHP-GDM by allowing more flexibility in the specific problem under consideration. The Induced Ordered Weighted Average (IOWA) operator is a promising tool for decision making with the ability to reflect the complex attitudinal character of decision makers. The Maximum Entropy OWA (MEOWA) which is based on the maximum entropy principle and the level of 'orness' is a systematic way to derive weights for decision analysis. In this paper, the focus is given on the integration of some induced aggregation operators with the AHP-GDM based-MEOWA as an extension model. An illustrative example is presented to show the results obtained with different types of aggregation operators.

**Keywords:** AHP-GDM, Maximum Entropy OWA, Induced Generalized OWA.

## 1 Introduction

Multiple Criteria Decision Making (MCDM) is one of the active topics in Operations Research. In general, MCDM can be considered as a process of selecting one alternative from a set of discrete alternatives with respect to several criteria. One of the most widely used MCDM techniques is the Analytic Hierarchy Process (AHP), which was proposed by Saaty (Saaty, 1977). The AHP is based on the judgment of problems with multiple criteria by means of the construction of a ratio scale corresponding to the priorities of alternatives. Since its introduction, AHP has been used in many applications. More details on the analysis and review of AHP and its applications can be referred for instance in (Vaidya and Kumar, 2006; Saaty, 2013). In order to deal with complex decision making problems involving multiple experts' assessments, the AHP has been extended to AHP-Group Decision Making (AHP-GDM) model. Escobar and Moreno-Jimerez (2007) and Gargallo et al. (2007) for example, have modified the conventional AHP method to AHP-GDM and showed the effectiveness and reliability

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 476–485, 2014.

of the model. The AHP are popularly used in applications due to some advantages, include a hierarchy structure by reducing multiple criteria into a pairwise comparison method for individual or group decision making and it allows the use of quantitative and qualitative information in evaluation process.

The Ordered Weighted Average (OWA) on the other hand is a family of aggregation procedures and was developed by Yager (Yager, 1988). The OWA operators are introduced to provide a parameterized class of mean-type aggregation operators. These include a family of aggregation operators that lie between the 'and' (min) and the 'or' (max), and a unified framework for decision making under uncertainty. Subsequently, Yager (2004) generalizes the OWA operator by combining it with other mean operators and called it the Generalized OWA (GOWA). This includes a wide range of mean operators, such as the OWA, Ordered Weighted Geometric Average (OWGA), Ordered Weighted Harmonic Average (OWHA) and Ordered Weighted Quadratic Average (OWQA) operators. The Induced OWA (IOWA) is another extension of the OWA operator. Yager and Filev (1999) inspired by the work of (Mitchell and Estrakh, 1997) have introduced the IOWA operator as a general type of aggregation. The main difference between OWA and IOWA is in the reordering step. Instead of directly order the argument values as in OWA operator, the IOWA operator utilizes another mechanism called induced variables as a pair of argument values, in which it uses a first component to induce the argument values of a second component. The advantage of IOWA is the ability to consider complex attitudinal character of the decision makers that provide a complex picture of the decision making process. The IOWA operator has been studied by a lot of authors in recent years, see for example (Chiclana et al. 2004; Xu, 2006; Merigo and Casanovas, 2011).

Recently, much attention has been given on the integration of AHP with OWA operator (i.e., concentrated on aggregation process) as inspired by the work of (Yager and Kelman, 1999). At first, Yager and Kelman (1999) have proposed the extension of AHP using OWA operator with fuzzy linguistic quantifier. This approach generalizes the weighted average normally used in AHP to OWA based linguistic quantifier. In addition to OWA based fuzzy linguistic quantifier technique, the Maximum Entropy OWA (MEOWA) operator has been proposed to be used in decision making analysis. O' Hagan (1988, 1990) has developed a maximum entropy approach, which formulates the problem as a constraint nonlinear optimization model with a predefined degree of orness as constraint and entropy as objective function, and used it to determine OWA weights. Subsequently, Filev and Yager (1995) have examined the analytical properties of MEOWA operators and proposed the analytic approach of MEOWA operator. Since that the MEOWA has been used in many applications, include in MCDM area (Yager, 2009; Ahn, 2011).

In this paper, the focus is given on the integration of induced aggregation operators with MEOWA weights in AHP-GDM as an extension model. The reason for doing this is because there are situations in which it is necessary to aggregate the variables with an inducing order instead of aggregating with the conventional OWA operator. For example, such a method is useful when the attitudinal character of the decision maker is particularly complex or when there are a number of external factors (i.e., personal effects on each alternative) affecting the decision analysis. The general framework model for AHP-GDM which include different types of aggregation operators is proposed. The main advantage of this approach is the possibility of considering

a wide range of induced aggregation operators. Therefore, the decision makers get a more complete view of the problem and able to select the alternative that it is in accordance with their interests. These problems are studied in detail by conducting an extensive analysis of some families of induced aggregation operators.

The remainder of this paper is organized as follows. In Section 2, the different types of aggregation operators are reviewed in general. Section 3 briefly discusses the AHP method. Section 4 examines the MEOWA technique. In section 5, the process of using the IGOWA operators in AHP-GDM is discussed. Section 6 provides an illustrative example of the developed method and Section 7, the conclusions of the paper.

## 2 Preliminaries

In the following, the basic aggregation operators that are used in this paper are briefly discussed.

*Definition 1* (Yager, 1988). An OWA operator of dimension $n$ is a mapping $OWA: R^n \to R$ that has an associated weighting vector $W$ of dimension $n$ such that $\sum_{j=1}^{n} w_j = 1$ and $w_j \in [0,1]$, then:

$$OWA(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j \tag{1}$$

where $b_j$ is the $j$th largest $a_i$ and $R$ is the set of positive real numbers.

*Definition 2* (Yager and Filev, 1999). An IOWA operator of dimension $n$ is a mapping $IOWA: R^n \to R$ that has an associated weighting vector $W$ of dimension $n$ such that $\sum_{j=1}^{n} w_j = 1$ and $w_j \in [0,1]$, then:

$$IOWA(\langle u_1, a_1\rangle, \langle u_2, a_2\rangle, \ldots, \langle u_n, a_n\rangle) = \sum_{j=1}^{n} w_j b_j \tag{2}$$

where $b_j$ is the $a_i$ value of the IOWA pair $\langle u_i, a_i\rangle$ having the $j$th largest $u_i$, $u_i$ is the order-inducing variable and $a_i$ is the argument variable. Note that, in case of 'ties' between argument values, the policy proposed by Yager and Filev (1999) will be implemented, in which each argument of tied IOWA pair is replaced by their average.

*Definition 3* (Merigo and Gil-Lafuente, 2009). An IGOWA operator of dimension $n$ is a mapping $IGOWA: R^n \to R$ that has an associated weighting vector $W$ of dimension $n$ such that $\sum_{j=1}^{n} w_j = 1$ and $w_j \in [0,1]$, then:

$$IGOWA(\langle u_1, a_1\rangle, \langle u_2, a_2\rangle, \ldots, \langle u_n, a_n\rangle) = \left(\sum_{j=1}^{n} w_j b_j^{\lambda}\right)^{1/\lambda} \tag{3}$$

where $b_j$ is the $a_i$ value of the IGOWA pair $\langle u_i a_i\rangle$ having the $j$th largest $u_i$, $u_i$ is the order inducing variable, $a_i$ is the argument variable and $\lambda$ is a parameter such that $\lambda \in (-\infty, \infty)$. With different values of $\lambda$, various type of weighted average can be derived. For instance, when $\lambda = -1$, IOWHA operator can be derived, when $\lambda = 2$, the IOWQA operator is derived. The OWA, the IOWA, and the IGOWA operators are all commutative, monotonic, bounded and idempotent.

## 3 The Analytical Hierarchy Process Method

The AHP is introduced based on the weighted average model for complex decision making problems (Saaty, 1977; Saaty, 2013) or also known as multiplicative preference relation. The AHP can be divided into three major steps; developing the AHP hierarchy, pairwise comparison of elements of the hierarchical structure and constructing an overall priority rating. Specifically, the pairwise comparison matrix for each level has the following form: Let $x_i = [x_{ij}]_{m\times m}$ where $x_{ij}$ is pairwise comparison rating for components $i$ and components $j$ $(i, j = 1,2, \dots, m)$. The matrix $x_i$ is reciprocal, such that $x_{ij} = {x_{ji}}^{-1}$ for $(i \neq j)$ and all its diagonal elements are unity, $x_{ij} = 1$, $(i = j)$. In order to measure the degree of consistency, calculate the consistency index ($CI$) as follows:

$$CI = \frac{\lambda_{max} - p}{p-1} \tag{4}$$

where $\lambda_{max}$ is the biggest eigenvalue that can be obtained once we have its associated eigenvector and $m$ is the number of columns of matrix $X$. Further, we can calculate the consistency ratio ($CR$), which is defined as follows:

$$CR = \frac{CI}{RI} \tag{5}$$

where $RI$ is the random index, the consistency index of a randomly generated pairwise comparison matrix. It can be shown that $RI$ depends on the number of elements being compared. The table for $RI$ can be referred in Saaty (1977). The consistency ratio ($CR$) is designed in such a way that if $CR < 0.10$ then the ratio indicates a reasonable level of consistency in the pairwise comparison.

For the given hierarchical structure, the overall evaluation score, $D_i$ of the ith alternative is calculated as follows: $AHP\ (D_i) = \sum_{j=1}^{n} w_j x_{ij}$. The performance of alternatives $X_i$ with respect to criteria $c_j$ is described by a set of criteria values $X_i = [x_{ij}]$; $x_{ij} \in [0,1]$ for $i = 1,2, \dots, m$ and $j = 1,2, \dots, n$. The evaluation process in the AHP uses a simple weighted linear combination to calculate the local scores of each alternative.

## 4 Maximum Entropy OWA

Various approaches have been suggested for obtaining the weights in decision making process. Motivated by the maximum entropy principle, O'Hagan (1988, 1990) has developed a way to generate OWA weights by maximizing the entropy which subject to the weight constraint and the value of the attitudinal character (or degree of orness). The methodology is based on the mathematical programming problem and have come to be known as the Maximum Entropy OWA weights. It can be noticed that MEOWA weights used to spread the weights as uniformly as possible and at the same time satisfying the attitudinal character.

Filev and Yager (1995) obtained an analytic solution for the determination of the MEOWA weights. In particular, the authors showed that the MEOWA weights for an aggregation of degree n can be expressed as:

$$v_j = \frac{e^{\beta^* \times ((n-i)/(n-1))}}{\sum_{j=1}^{n} e^{\beta^* \times ((n-i)/(n-1))}}, j = 1,2,\dots,n \tag{6}$$

where $\beta^* \in (-\infty, \infty)$ is a parameter dependent upon the value $\alpha$, which is desired attitudinal character. Specifically, they showed that $\beta^* = (n-1)\ln(h)$, where $h^*$ as a positive solution of the equation $\sum_{j=1}^{n}[(n-j)/(n-1) - \alpha]h^{(n-j)} = 0$.

In what follows, the MEOWA weight will be used to be integrated with AHP-GDM in the next section. In this case, the value $\alpha$ is computed based on the weight of AHP pairwise comparison (criteria weight or relative importance of criteria). The attitude of expert in differentiate each criterion under consideration determine the degree of orness. Then, based on the degree of orness, MEOWA weight, or defined as ordered weight can be calculated.

## 5 An Extension of the AHP-Group Decision Making Method

In this section, an extension of the AHP method under group decision making is presented. In what follows, the proposed method is represented step by step as in the consequence algorithm. Assume $x_i \in X, (x_i, i = 1,2,\dots,m)$ comprise a finite set of alternatives. Let $c_j \in C, (c_j, j = 1,2,\dots,n)$ and $c_k^j \in C$, $(c_k^j, j = 1,2,\dots,n;\ k = 1,2,\dots,p)$ are the criteria and sub-criteria under consideration, respectively. Then, let $E_h \in E$, $(E_h, h = 1,2,\dots,r)$, be a group of experts, with each expert $E_h$ presenting his/her preferences or opinions for rating the alternatives $x_i$, and weighting the criteria $c_j$ (or sub-criteria $c_k^j$). Based on the above concepts, the algorithm for the IGOWA AHP-GDM consists of the following steps.

*Step 1*: Each decision maker or expert $(E_h, h = 1,2,\dots,r)$, compares the $m$ alternatives $(x_i, i = 1,2,\dots,m)$ and provides a pairwise comparison matrix:

$$x_i^{[h]} = \left[x_{ij}\right]_{m\times m}^{[h]}, i,j = 1,2,\dots,m \tag{7}$$

with $(\forall i = j,\ x_{ij} = 1)$ and $(\forall i \neq j,\ x_{ij} = \left[x_{ji}\right]^{-1})$. Then, the alternatives values $x_i^{[h]}$ for each expert $E_h$ can be calculated as follow:

$$x_i^{[h]} = \sum_{i=1}^{m} x_{ij}^{[h]} \times \left[\sum_{i=1}^{m}\sum_{j=1}^{m} x_{ij}^{[h]}\right]^{-1}, i,j = 1,2,\dots,m \tag{8}$$

*Step 2*: Compute the judgment matrix for a group of experts $x_i^{[G]}$. Consider $\Theta_h$ as the weight that the $h$th expert $(E_h, h = 1,2,\dots,r)$ has in forming the group decision $(\Theta_h \geq 0;\ \sum_{h=1}^{r}\Theta_h = 1)$.

$$x_i^{[G]} = \prod_{h=1}^{r}(x_i^{[h]})^{\Theta_h}, i = 1,2,\dots,m \tag{9}$$

*Step 3*: Calculate the pairwise comparison matrix for the criteria $c_j \in C, (c_j,\ j = 1,2,\dots,n)$ and sub-criteria $c_k^j \in C, (c_k^j, j = 1,2,\dots,n; k = 1,2,\dots,p)$. Then derive the criteria weights $\omega_j$ and sub-criteria weights $\omega_k^j$ using the same formulation as for alternatives, equations (8) and (9).

*Step 4*: Calculate the composite weights of the criteria $c_j$ and sub-criteria $c_k^j$.

$$w_j = \omega_j \times \omega_k^j \tag{10}$$

*Step 5*: Compute the orness value $\alpha(w_j)$ using the Maximum Entropy OWA to calculate the ordered weight, $v_j^i$.

$$\alpha(w_j) = \left(\sum_{j=1}^{n}(n-j)u_j\right)/(n-1) \tag{11}$$

where $u_j$ is the reordered $j$th criteria weight, $w_j$ (or composite weight) associated with $x_i^{[G]}$.

*Step 6*: Find a positive solution $h^*$ of the algebraic equation.

$$\sum_{i=1}^{n}[(n-i)/(n-1) - \alpha]h^{(n-i)} = 0 \tag{12}$$

*Step 7*: Compute the ordered weight $v_j$ using the equation (6), where $0 \le v_j \le 1$ and $\sum_{j=1}^{n} v_j = 1$.

*Step 8*: Finally, the overall score $D_i(x)$ of the ith alternative is defined as the summation of the product of weight of each criterion by the performance of the alternative with respect to that criterion,

$$IGOWA - AHP - GDM\ [D_i(x)] = \left(\sum_{j=1}^{n} v_j^i z_j^{\lambda}\right)^{\lambda} \tag{13}$$

where $z_{ij}$ is the $x_i^{[G]}$ value of the OWA pair $\langle t_i, x_i^{[G]}\rangle$ having the $j$th largest $t_i$ of the order-inducing variable, and $x_i^{[G]}$ is the argument variable. Besides, $v_j$ is the ordered weight based on Maximum Entropy OWA and $\lambda$ is a parameter such that $\lambda \in (-\infty, \infty)$, with different values of $\lambda$ reflect various types of weighted average.

Note that, when $t_i = x_i^{[G]}$, the IGOWA-AHP-GDM turn to GOWA-AHP-GDM. Similarly, when $x_i^{[G]}$ values are not arranged using OWA function, then the method turn to conventional AHP-GDM.

## 6 Illustrative Example

An illustrative example is given to implement the methodologies discussed in the previous sections. For this purpose, let consider an investment selection problem where a company is looking for an optimal investment. There are five possible alternatives to be considered as follows: $x_1$ is a computer company; $x_2$ is a chemical company; $x_3$ is a food company; $x_4$ is a car company; $x_5$ is a TV company.

In order to evaluate these alternatives, a group of experts must make a decision according to the following four attributes: $C_1$ = risk analysis; $C_2$ = growth analysis; $C_3$ = social-political impact analysis; and $C_4$ = environmental impact analysis.

In this case, assume that three experts involved and the weight vector for the experts is $(\Theta_h = (0.5,0.3,0.2), h = 1,2,3$. Due to the fact that the attitudinal character is very complex because it involves the opinion of different members of the board of directors, the experts use order inducing variables to represent it as shown in Table 1.

**Table 1.** Inducing variables

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| $X_1$ | 25 | 18 | 24 | 16 |
| $X_2$ | 12 | 34 | 18 | 22 |
| $X_3$ | 22 | 13 | 28 | 21 |
| $X_4$ | 31 | 24 | 14 | 20 |
| $X_5$ | 30 | 25 | 23 | 16 |

First, let all the experts agreed to provide pairwise comparison of criteria as shown in Table 2. In this case, no sub-criteria considered. Hence, based on criteria weight technique, the weight for each criterion can be derived and consistency ratio is then computed to check the consistency of pairwise comparison.

**Table 2.** Pairwise comparison matrix and the weight ratio of criteria

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $w_j$ | |
|---|---|---|---|---|---|---|
| $C_1$ | 1 | 0.5 | 2 | 4 | 0.3111 | *CR*=0.036 |
| $C_2$ | 2 | 1 | 2 | 3 | 0.4064 | |
| $C_3$ | 0.5 | 0.5 | 1 | 2 | 0.1824 | |
| $C_4$ | 0.25 | 0.3333 | 0.5 | 1 | 0.1001 | |

Subsequently, the weights $w_j$ are proceed to be calculated with the MEOWA. The results of this weight are presented in Table 3, where α is the measure of orness, $h^*$ is the positive solution of algebraic equation, $\beta^*$is the value that relates to the weights and the measure of orness, and $v_j$ is the weight of MEOWA.

**Table 3.** The α, $h^*$, $\beta^*$ values of criteria weights and MEOWA weights $v_j$

| | α | $h^*$ | $\beta^*$ | $v_1$ | $v_2$ | $v_3$ | $v_4$ |
|---|---|---|---|---|---|---|---|
| $X_1$ | 0.5682 | 1.1793 | 0.4948 | 0.3148 | 0.2669 | 0.2263 | 0.1919 |
| $X_2$ | 0.5339 | 1.0849 | 0.2445 | 0.2813 | 0.2593 | 0.2390 | 0.2203 |
| $X_3$ | 0.4232 | 0.8301 | -0.5586 | 0.1850 | 0.2229 | 0.2685 | 0.3235 |
| $X_4$ | 0.6154 | 1.3275 | 0.8499 | 0.3639 | 0.2741 | 0.2065 | 0.1555 |
| $X_5$ | 0.6429 | 1.4262 | 1.0650 | 0.3941 | 0.2763 | 0.1937 | 0.1358 |

Next, each expert provides rating (or pairwise comparison) for all alternatives with respect to each criterion in order to get relative performance of alternatives in specific criterion. The results of standardized performance of each expert are shown in Tables 4, 5 and 6, respectively.

**Table 4.** Standardized performance ratings of Expert 1

| $E_1$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| $X_1$ | 0.1866 | 0.2412 | 0.2618 | 0.2571 |
| $X_2$ | 0.3069 | 0.1353 | 0.0892 | 0.0881 |
| $X_3$ | 0.0573 | 0.0743 | 0.1528 | 0.1539 |
| $X_4$ | 0.3069 | 0.1353 | 0.0526 | 0.4129 |
| $X_5$ | 0.1422 | 0.4137 | 0.4436 | 0.0881 |

**Table 5.** Standardized performance ratings of Expert 2

| $E_2$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| $X_1$ | 0.0881 | 0.1618 | 0.0604 | 0.0890 |
| $X_2$ | 0.4129 | 0.2760 | 0.1382 | 0.1579 |
| $X_3$ | 0.0881 | 0.1054 | 0.3972 | 0.2976 |
| $X_4$ | 0.2571 | 0.0596 | 0.0954 | 0.2976 |
| $X_5$ | 0.1539 | 0.3971 | 0.3088 | 0.1579 |

**Table 6.** Standardized performance ratings of Expert 3

| $E_3$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| $X_1$ | 0.0890 | 0.1042 | 0.0743 | 0.0986 |
| $X_2$ | 0.2976 | 0.3902 | 0.1353 | 0.1611 |
| $X_3$ | 0.1579 | 0.0588 | 0.2412 | 0.4162 |
| $X_4$ | 0.2976 | 0.1505 | 0.1353 | 0.0624 |
| $X_5$ | 0.1579 | 0.2962 | 0.4137 | 0.2618 |

Then, the results for each expert can be aggregated to form a matrix for group consensus. Table 7 presented the aggregated performance of experts.

**Table 7.** Aggregated performance of experts

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| $X_1$ | 0.1285 | 0.1809 | 0.1311 | 0.1544 |
| $X_2$ | 0.3334 | 0.2071 | 0.1105 | 0.1184 |
| $X_3$ | 0.0798 | 0.0788 | 0.2230 | 0.2288 |
| $X_4$ | 0.2893 | 0.1081 | 0.0760 | 0.2565 |
| $X_5$ | 0.1487 | 0.3823 | 0.3924 | 0.1305 |

Finally, the overall score $D_i(x)$ of each alternative ith is derived as the summation of the product of MEOWA weights by the aggregated performance of experts. With this information, different results are obtained using different types of IGOWA operators. The final results of the aggregation process with different operators are shown in Tables 8 and 9. Meanwhile the ordering of investments is shown in Table 10.

**Table 8.** Aggregated results 1

| AM | WA | OWA | OWHA | OWQA | IOWA |
|---|---|---|---|---|---|
| 0.149 | 0.147 | 0.151 | 0.148 | 0.152 | 0.146 |
| 0.192 | 0.226 | 0.220 | 0.180 | 0.239 | 0.189 |
| 0.153 | 0.128 | 0.122 | 0.098 | 0.139 | 0.146 |
| 0.182 | 0.193 | 0.186 | 0.137 | 0.207 | 0.200 |
| 0.263 | 0.258 | 0.281 | 0.219 | 0.307 | 0.258 |

**Table 9.** Aggregated results 2

| IOWHA | IOWQA | GM | WG | OWG | IOWG |
|---|---|---|---|---|---|
| 0.143 | 0.148 | 0.147 | 0.145 | 0.139 | 0.145 |
| 0.157 | 0.208 | 0.173 | 0.205 | 0.199 | 0.171 |
| 0.112 | 0.163 | 0.134 | 0.112 | 0.108 | 0.128 |
| 0.150 | 0.219 | 0.157 | 0.167 | 0.160 | 0.175 |
| 0.204 | 0.285 | 0.232 | 0.229 | 0.250 | 0.229 |

**Table 10.** Ranking of the investments

| | Ranking | | Ranking |
|---|---|---|---|
| *AM* | $A_5 \succ A_2 \succ A_4 \succ A_3 \succ A_1$ | *IOWHA* | $A_5 \succ A_2 \succ A_4 \succ A_1 \succ A_3$ |
| *WA* | $A_5 \succ A_2 \succ A_4 \succ A_1 \succ A_3$ | *IOWQA* | $A_5 \succ A_4 \succ A_2 \succ A_3 \succ A_1$ |
| *OWA* | $A_5 \succ A_2 \succ A_4 \succ A_1 \succ A_3$ | *GM* | $A_5 \succ A_2 \succ A_4 \succ A_1 \succ A_3$ |
| *OWHA* | $A_5 \succ A_2 \succ A_1 \succ A_4 \succ A_3$ | *WG* | $A_5 \succ A_2 \succ A_4 \succ A_1 \succ A_3$ |
| *OWQA* | $A_5 \succ A_2 \succ A_4 \succ A_1 \succ A_3$ | *OWG* | $A_5 \succ A_2 \succ A_4 \succ A_1 \succ A_3$ |
| *IOWA* | $A_5 \succ A_4 \succ A_2 \succ A_3 = A_1$ | *IOWG* | $A_5 \succ A_4 \succ A_2 \succ A_1 \succ A_3$ |

## 7 Conclusions

This paper has presented an extension of the Analytical Hierarchy Process method under Group Decision Making (AHP-GDM) with some induced aggregation operators. The Maximum Entropy OWA (MEOWA) has been proposed to derive weights in the AHP-GDM model. First, some modifications have been made to generalize the aggregation process used in AHP-GDM with some Induced Generalized Ordered Weighted Average (IGOWA) operators. The main advantages of this approach are the ability to deal with the complex attitudinal character of the decision makers and the aggregation of the information with a particular reordering process. Therefore, the decision makers get a more complete view of the problem and able to select the alternative that it is in accordance with their interests. The procedure of the AHP-GDM method with IGOWA operators has been discussed in detail. A numerical example on investment selection problem has been given to exemplify the feasibility of the proposed method. The comparison of some induced aggregation operators has also been made.

## References


1. Chiclana, F., Herrera-Viedma, E., Herrera, F., Alonso, S.: Induced ordered weighted geometric operators and their use in the aggregation of multiplicative preference relations. International Journal of Intelligent Systems 19, 233–255 (2004)
2. Escobar, M.T., Moreno-Jimenez, J.M.: Aggregation of individual preference structures in AHP-Group Decision Making. Group Decision and Negotiation 16, 287–301 (2007)
3. Filev, D.P., Yager, R.R.: Analytic properties of Maximum Entropy OWA operators. Information Sciences 85, 11–27 (1995)
4. Gargallo, P., Moreno-Jimenez, J.M., Salvador, M.: AHP-Group Decision Making: A Bayesian approach based on mixtures for group pattern identification. Group Decision and Negotiation 16, 485–506 (2007)
5. Merigo, J.M., Casanovas, M.: Induced aggregation operators in the Euclidean distance and its application in financial decision making. Exp. Syst. Appl. 38, 7603–7608 (2011)
6. Mitchell, H.B., Estrakh, D.D.: A modified OWA operator and its use in lossless DPCM image compression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 5, 429–436 (1997)
7. O'Hagan, M.: Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic. In: IEEE Asilomar Conference on Signals, Systems and Computers, pp. 681–689 (1988)
8. O'Hagan, M.: A fuzzy neuron based upon maximum entropy-ordered weighted averaging. In: Bouchon-Meunier, B., Zadeh, L.A., Yager, R.R. (eds.) IPMU 1990. LNCS, vol. 521, pp. 598–609. Springer, Heidelberg (1991)
9. Saaty, T.L.: A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology 15, 234–281 (1977)
10. Saaty, T.L.: The modern science of multicriteria decision making and its practical applications: The AHP/ANP Approach. Operations Research 61, 1101–1118 (2013)
11. Vaidya, O.S., Kumar, S.: Analytic hierarchy process: An overview of applications. European Journal of Operational Research 169, 1–29 (2006)
12. Xu, Z.S.: Induced uncertain linguistic OWA operators applied to group decision making. Information Fusion 7, 231–238 (2006)
13. Yager, R.R., Kelman, A.: An extension of the analytical hierarchy process using OWA operators. Journal of Intelligent and Fuzzy Systems 7, 401–417 (1999)
14. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems and Man Cybernetics 18, 183–190 (1988)
15. Yager, R.R.: Generalized OWA aggregation operators. Fuzzy Optimization and Decision Making 3, 93–107 (2004)
16. Yager, R.R., Filev, D.P.: Induced Ordered Weighted Averaging Operators. IEEE Transactions on Systems and Man Cybernetics - Part B 29, 141–150 (1999)
17. Ahn, B.S.: Compatible weighting method with rank order centroid: Maximum entropy ordered weighted averaging approach. European Journal of Operational Research 212, 552–559 (2011)

# Calibration of Utility Function and Mixture Premium

Jana Špirková

Faculty of Economics, Matej Bel University,
Tajovského 10, 975 90 Banská Bystrica, Slovakia

**Abstract.** Calibration of the different types of utility functions of money is discussed in this paper. This calibration is based on an expected utility maximization of different alternatives of investment strategies which are offered to persons in a short questionnaire. Investigated utility functions have different Arrow-Pratt absolute and relative risk aversion coefficients. Moreover, the paper proposes a basic concept of uncertainty modeling by the chosen utility functions for the determination of so-called maximum mixture premium in non-life insurance. This concept is based on aggregation of maximum premiums by so-called mixture function with the selected weighting function. A case study is included.

**Keywords:** utility, calibration, risk aversion, mixture premium.

## 1 Introduction

The aim of this paper is to propose a new possibility of determining the maximum gross annual premium in non-life insurance. The paper focuses on the modeling of uncertainty by calibration of two chosen personal utility functions of money, by expected utility maximization and by aggregation. More precisely, the paper proposes a model for determination of maximum premium acceptable for both the insured and the insurer.

Uncertainty modeling analysis investigates the uncertainty of variables that are used in decision-making problems. In other words, uncertainty analysis aims to make a technical contribution to decision-making through the quantification of uncertainties in the relevant variables. Utility function may be used as a basis to describe individual approaches to risk. In general, there are three basic approaches to risk. Two of them refer to *risk loving* and *risk averse*, which are accepted only in gambling. And the third one, *risk-neutral*, is between these two extremes. Risk-neutral behavior is typical of persons who are enormously wealthy. Many people may be both risk averse and risk loving, depending on the range of monetary values being considered.

A more modern approach of the utility theory was advanced by John von Neumann and Oskar Morgenstern in 1947 in their book *Theory of Games and Economic Behavior* [11]. In 2007, the 60th-anniversary edition of this book was published [12]. These authors proposed that a utility function may be tailored for any individual, under certain conditions. These conditions provide several valid,

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 486–495, 2014.

basic shapes for the utility function. We can find a very interesting approach about utility functions in [8], [9], [13] and [15].

Our paper has 5 sections. In Section 2, we give preliminaries of basic properties of utility functions. Moreover, we mention possibilities of averaging aggregation by so-called mixture operators (mixture functions). In Section 3, we give a possible way of calibration of utility functions by a short personal questionnaire and by the criterion of expected utility maximization. Section 4 represents a case study - the application of calibrated utility functions on modeling of maximum premium in non-life insurance. Moreover, it contains a concept of so-called mixture premium, which represents suitable aggregation of multiple premium values determined on the basis of calibrated utility functions, that means on the basis of different approaches to risk. Finally, Section 5 contains conclusions and indications of our next investigation of the mentioned topic.

## 2 Preliminaries

This paper focuses on two selected utility functions with different properties, specifically, with different absolute and relative risk aversion coefficients, and moreover, with different values of expected utilities in sign. At the beginning, we give the basic properties of these functions.

### 2.1 Utility Function and Expected Utility

The first investigated utility function is exponential and it is given by

$$u(x) = \frac{1}{\alpha} \cdot \left(1 - exp^{-\alpha \cdot x}\right) \quad if\ \alpha > 0. \tag{1}$$

The second one is a power function and it can be written as follows:

$$u(x) = \frac{x^{1-\alpha}}{1-\alpha} \ \ if\, \alpha > 1. \tag{2}$$

Both above-mentioned functions represent a standard class of utility functions, but with different properties regarding Arrow-Pratt absolute and relative risk aversion coefficients. While the function (1) has a constant absolute risk aversion coefficient ($ARA$)

$$ARA(x)_{exp} = -\frac{u''(x)}{u'(x)} = \alpha, \tag{3}$$

the function (2) has not, and it is given as follows

$$ARA(x)_{power} = \frac{\alpha}{x}. \tag{4}$$

The function (1) is characterized by a relative risk aversion coefficient ($RRA$)

$$RRA(x)_{exp} = -x \cdot \frac{u''(x)}{u'(x)} = x \cdot \alpha, \tag{5}$$

while the function (2) by a constant, and it is given by

$$RRA(x)_{power} = \alpha. \tag{6}$$

The well-known theorem below describes properties of the utility function and its expected utility. It can written as follows.

**Theorem 1.** *(Jensen's inequality) [13] Let $X$ be a random variable (with a finite expectation). Then,*
*if $u(x)$ is concave,*

$$E\left[u\left(X\right)\right] \leq u\left(E\left[X\right]\right). \tag{7}$$

*If $u(x)$ is convex,*

$$E\left[u\left(X\right)\right] \geq u\left(E\left[X\right]\right). \tag{8}$$

*Equality holds if and only if $u\left(x\right)$ is linear according to $X$ or $var\left(X\right) = 0$.*

### 2.2 Expected Utility Maximization

Expected utility is calculated by the well-known formula

$$E\left[u\left(X\right)\right] = \sum_{i=1}^{n} u(x_i) \cdot p_i, \tag{9}$$

where $X = (x_1, x_2, \ldots, x_n)$ is a vector of the possible alternatives and $p_i$ is the probability of alternative $x_i$ , $i = 1, 2, \cdots, n$.

The criterion of expected utility maximization corresponds to the preference order $\preceq$, for which

$$X \preceq Y \equiv E\left[u(X)\right] \preceq E\left[u(Y)\right] \tag{10}$$

for a utility function $u$. The relation (10) means that among two random variables $X$, $Y$ we prefer the random variable with the larger expected utility. If $u(x)$ is non-decreasing, the rule (10) is monotone. The investor who follows (10) is called an expected utility maximizer.

### 2.3 Mixture Function

In many applications we can meet with averaging aggregation of a set of inputs into one representative value. There is a growing interest in aggregation functions, which include, under certain conditions, also mixture functions. Books [2] and [7] bring an overview of almost all aggregation methods.

The application in this paper focuses on aggregation of selected maximum premiums by a mixture function. We stress that the weights of this function are functions of input values.

Let $[a, b] \subset R$ be a closed non-trivial interval and $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be an input vector. Then, the mixture function can be defined as follows.

**Definition 1.** *The mapping* $M_g : [a, b]^n \to [a,\ b]$ *with a weighting differentiable function* $g : [a, b]^n \to [0,\ \infty[$ *given by*

$$M_g(x_1, \ldots, x_n) = \frac{\sum_{i=1}^{n} g(x_i) \cdot x_i}{\sum_{i=1}^{n} g(x_i)} \tag{11}$$

*is called a mixture operator.*

We have to emphasize that in general, mixture functions need not be monotone non-decreasing (monotone non-decreasingness is one of basic properties of aggregation functions). There exist sufficient conditions for their monotonicity. For the purposes of our paper, we give here only two sufficient conditions for non-decreasingness of a mixture function with non-increasing weighting function $g$ which are given by $g(x) + g'(x) \geq 0$ or $g(x) + g'(x) \cdot x \geq 0$ . Other sufficient conditions can be found, for example in [7] or [14].

## 3 Calibration of Utility Function

The selected utility functions (1) and (2) which are mentioned in Section 2, are created by the coefficient $\alpha$, which in general represents the approach to risk. For the purpose of calibration of the mentioned utility functions we have designed the following questionnaire because it is very clear and time-saving, [10].
Suppose that you are going to invest 17,000 euros and you have a choice between four different investment strategies for a three-year investment. Which of these alternatives would you prefer?

- Alternative $A_1$: in the best case profit 1,700 euros (10.00 %)[1], in the worst case profit 550 euros (3.24 %).
- Alternative $A_2$: in the best case profit 2,600 euros (15.29 %), in the worst case zero profit (but no loss), (0.00 %).
- Alternative $A_3$: in the best case profit 4,000 euros (23.53 %), in the worst case loss 1,700 euros (-10.00 %).
- Alternative $A_4$: in the best case profit 6,500 euros (38.53 %), in the worst case loss 4,000 euros (-23.53 %).

Respondents have chosen one of the previous alternatives. It is apparent that individual alternatives are put in order, so that the first alternative is of the lowest risk and the fourth alternative is of the highest risk. We evaluated expected utilities for all alternatives and for some selected risk aversion coefficients $\alpha$ in the utility functions (1) and (2). On the basis of the equation (9) we evaluated expected utilities and determined maximal expected utilities by

$$E[u(x)] = p \cdot u(x_1) + (1 - p) \cdot u(x_2) \to max \tag{12}$$

[1] Percentage of possible profit or loss sequentially in all alternatives.

for all alternatives on the level of all used $\alpha$. Our results are written in Tab. 1 and Tab. 2. Observe, that all values - the best and the worst cases we divided by $10,000$ for easier evaluation with respect to suitable decimal places.

Tab. 1 and Tab. 2 show the expected utilities with respect to each alternative and coefficient $\alpha$. It is obvious the expected utilities depend on the probability of occurence of the best case $x_1$ or the worst case $x_2$. For each $\alpha$, we look for an alternative with the maximum expected utility. For each alternative, there is a set of $\alpha$s. From this set we take the $\alpha$ with the highest expected utility. For alternatives $A_2, A_3$ and $A_4$, we set the probability at $p = 0.20$ and $1 - p = 0.80$, respectively. If we set the same probability for alternative $A_1$, we would obtain an empty set of $\alpha$s. In order to show the essence of our calibration, for $A_1$ we have chosen the probability $p = 0.99$ and $1 - p = 0.01$. The results in Tab. 1 show that the maximum expected utilities are sequential with $\alpha = 6, 3, 2, 1$. This means that the respondent who chooses alternative $A_1$ has utility function $u_1(x) = \frac{1}{6} \cdot \left(1 - exp^{-6x}\right)$. Another respondent who will choose the alternative $A_2$ has the utility function $u_2(x) = \frac{1}{3} \cdot \left(1 - exp^{-3x}\right)$, for alternative $A_3$, the utility function is given by $u_3(x) = \frac{1}{2} \cdot \left(1 - exp^{-2x}\right)$, and finally, for alternative $A_4$ the utility function is given by $u_4(x) = 1 - exp^{-x}$.

**Table 1.** Expected utilities of individual alternatives w.r.t. $\alpha$ and (1)

| Alternative | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| the best case $x_1$ | 1.87 | 1.96 | 2.10 | 2.35 |
| the worst case $x_2$ | 1.75 | 1.70 | 1.53 | 1.30 |
| probability $p$ | 0.99 | 0.80 | 0.80 | 0.80 |
| probability $(1-p)$ | 0.01 | 0.20 | 0.20 | 0.20 |
| | Expected utilities | | | |
| $\alpha = 1$ | 0.4568851 | 0.8507766 | 0.8587277 | **0.8691983** |
| $\alpha = 2$ | 0.8809222 | 0.4887262 | **0.4893129** | 0.4889345 |
| $\alpha = 3$ | 0.3321079 | **0.3321816** | 0.3321668 | 0,3317526 |
| $\alpha = 4$ | 0.2498581 | **0.2498656** | 0.2498451 | 0.2497076 |
| $\alpha = 5$ | 0.1999825 | **0.1999829** | 0.1999766 | 0.1999386 |
| $\alpha = 6$ | **0.1666644** | 0.1666644 | 0.1666629 | 0.1666530 |
| $\alpha = 7$ | **0.1428568** | 0.1428568 | 0.1428565 | 0.1428539 |
| $\alpha = 8$ | **0.1249999** | 0.1249999 | 0.1249999 | 0.1249992 |
| $\alpha = 9$ | **0.1111111** | 0.1111111 | 0.1111111 | 0.1111109 |
| $\alpha = 10$ | **0.0999999** | 0.0999999 | 0.0999999 | 0.0999999 |
| $\alpha = 11$ | **0.0909091** | 0.0909091 | 0.0909091 | 0.0909091 |
| $\alpha = 12$ | **0.0833333** | 0.0833333 | 0.0833333 | 0.0833333 |

Source: own construction

Tab. 2 gives expected utilities calculated by power function (2). The respondent who has chosen alternative $A_1$ has the utility function with risk aversion coefficient $\alpha = 12$, hence his/her utility function is given by $u(x) = \frac{-1}{11 \cdot x^{11}}$; the respondent who has chosen alternative $A_2$ will have $\alpha = 10$ and the corresponding utility function $u(x) = \frac{-1}{9 \cdot x^9}$; the respondent who has chosen alternative $A_3$

**Table 2.** Expected utilities of individual alternatives w.r.t. $\alpha$ and (2)

| Alternative | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|
| the best case $x_1$ | 1.87 | 1.96 | 2.10 | 2.35 |
| the worst case $x_2$ | 1.75 | 1.70 | 1.53 | 1.30 |
| probability $p$ | 0.99 | 0.80 | 0.80 | 0.80 |
| probability $(1-p)$ | 0.01 | 0.20 | 0.20 | 0.20 |
| | Expected utilities | | | |
| $\alpha = 2$ | -0.5351260 | -0.5258103 | -0.5116713 | **-0.4942717** |
| $\alpha = 3$ | -0.1431866 | -0.1387254 | -0.1334216 | **-0.1316026** |
| $\alpha = 4$ | -0.0510868 | -0.0489855 | **-0.0474083** | -0.0508922 |
| $\alpha = 5$ | -0.0205064 | -0.0195385 | **-0.0194081** | -0.0240641 |
| $\alpha = 6$ | -0.0087806 | **-0.0083486** | -0.0086885 | -0.0130056 |
| $\alpha = 7$ | -0.0039166 | **-0.0037327** | -0.0041531 | -0.0076975 |
| $\alpha = 8$ | -0.0017970 | **-0.0017247** | -0.0020903 | -0.0048420 |
| $\alpha = 9$ | -0.0008417 | **-0.0008175** | -0.0010969 | -0.0031722 |
| $\alpha = 10$ | -0.0004006 | **-0.0003956** | -0.0005955 | -0.0021362 |
| $\alpha = 11$ | **-0.0001930** | -0.0001948 | -0.0003324 | -0.0014663 |
| $\alpha = 12$ | **-0.0000939** | -0.0000974 | -0.0001898 | -0.0010205 |

Source: own construction

will have $\alpha = 5$, and the corresponding utility function $u(x) = \frac{-1}{4 \cdot x^4}$; and the respondent who has chosen alternative $A_4$ will have $\alpha = 3$ and the corresponding utility function $u(x) = \frac{-1}{2 \cdot x^2}$.

## 4 Personal Utility in Non-Life Insurance

In this section, we propose a model for determining the maximal annual gross premium in a non-life insurance. We use the utility functions in the shape (1) and (2).

### 4.1 Utility of the Insured

In general, our respondent has two alternatives - to buy insurance or not. Suppose that he/she owns a capital $w$, which he/she values wealth by the utility function $u$. If he/she is insured against a loss $X$ for a gross annual premium $GP$, he/she has a certain situation and his/her decision to buy insurance gives him/her the utility value $u\,(w - GP)$. If he/she is not insured, it means an uncertain situation for insured. In this case, the expected utility is $E\,[u\,(w - X)]$. Based on Jensen's inequality (7), we get

$$E\,[u\,(w - X)] \leq u\,(E\,[w - X]) = u\,(w - E\,[X]) \leq u\,(w - GP)\,. \qquad (13)$$

Since utility function $u$ is a non-decreasing continuous function, $GP \leq P^{max}$, where $P^{max}$ denotes the maximum premium to be paid. This so-called *zero utility premium* is the solution to the following utility equilibrium equation

$$E\,[u\,(w - X)] = u\,(w - P^{max})\,. \qquad (14)$$

On the basis of individual personal utility functions, we can determine maximum premium of our respondent - client of an insurance company will be willing to pay for insurance of his/her wealth on the basis of the following model: our client has 17,000 euros and he/she wants to insure his/her wealth worth 12,000 euros. The maximum premium $P^{max}$ is calculated by inverse function $u^{-1}$ to the utility equilibrium equation (14), which is given by

$$P^{max} = w - u^{-1}\left(E\left[u\left(w - X\right)\right]\right). \tag{15}$$

The previous function gives a maximum premium determined according to the exponential utility function (1) as follows

$$P^{max}_{exp} = \frac{ln(1 - p^* + p^* \cdot exp(\alpha x))}{\alpha}, \tag{16}$$

and according to the power function (2) by

$$P^{max}_{power} = w - \left((w - X)^{1-\alpha} \cdot p^* + w^{1-\alpha} \cdot (1 - p^*)\right)^{1/(1-\alpha)}, \tag{17}$$

where $p^*$ is the probability of occurence of insured event. In the case of our investigated utility functions the resulting values of maximum premiums with respect to risk aversion coefficients are given in Tab. 3 and Tab. 4, below. The first column in each table gives the probability of occurence of insured event $p^*$.

**Table 3.** Premiums according to utility function (1)

| probability $p^*$ | $E[X]$ | $P_1^{max}$ $\alpha = 6$ | $P_2^{max}$ $\alpha = 3$ | $P_3^{max}$ $\alpha = 2$ | $P_4^{max}$ $\alpha = 1$ | $MP_g$ |
|---|---|---|---|---|---|---|
| 0.001 | 12 | 1,415.80 | 116.60 | 49.87 | 23.17 | 53.99 |
| 0.002 | 24 | 2,170.10 | 229.25 | 99.24 | 46.30 | 107.31 |
| 0.003 | 36 | 2,687.48 | 338.23 | 148.13 | 69.36 | 159.95 |
| 0.004 | 48 | 3,081.74 | 443.75 | 196.55 | 92.38 | 211.98 |
| 0.005 | 60 | 3,400.33 | 546.04 | 244.50 | 115.34 | 263.40 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 0.900 | 10,800 | 11,824.54 | 11,658.90 | 11,523.34 | 11,552.15 | 11,652.00 |
| 1.000 | 12,000 | 12,000.00 | 12,000.00 | 12,000.00 | 12,000.00 | 12,000.00 |

Source: own construction

*Remark 1.* Note, that the insured is willing to pay a premium which is of equal value as the loss. For more information, see for example [4] and [16].

### 4.2 Mixture Premium

So far, we have examined individual alternatives and premiums only regardless of the number of respondents. Because we also have at hand a number of respondents who have chosen particular alternatives, we can aggregate maximum

**Table 4.** Premiums according to utility function (2)

| probability $p^*$ | $E[X]$ | $P_1^{max}$ $\alpha = 12$ | $P_2^{max}$ $\alpha = 10$ | $P_3^{max}$ $\alpha = 5$ | $P_4^{max}$ $\alpha = 3$ | $MP_g$ |
|---|---|---|---|---|---|---|
| 0.001 | 12 | 7,632.12 | 6,247.34 | 521.16 | 89.06 | 332.32 |
| 0.002 | 24 | 8,203.64 | 7,035.38 | 971.11 | 176.73 | 638.99 |
| 0.003 | 36 | 8,521.79 | 7,471.46 | 1,365.65 | 263.05 | 924.08 |
| 0.004 | 48 | 8,740.56 | 7,769.82 | 1,715.96 | 348.05 | 1,190.37 |
| 0.005 | 60 | 8,906.37 | 7,995.03 | 2,030.24 | 431.78 | 1,440.08 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 0.900 | 10,800 | 11,951.88 | 11,941.12 | 11,867.62 | 11,879.90 | 1,918.0 |
| 1.000 | 12,000 | 12,000.00 | 12,000.00 | 12,000.00 | 12,000.00 | 12,000.00 |

Source: own construction

premiums of all respondents, and thus to determine the most suitable maximum premium accepted by the insured and the insurer. We have 164 responses, namely 18 of the respondents ($n_1 = 18$) have chosen alternative $A_1$, 60 ($n_2 = 60$) have chosen $A_2$, 52 ($n_3 = 52$) have chosen $A_3$, and the remaining 34 respondents have chosen alternative $A_4$ ($n_4 = 34$).

Now, we propose a definition of so-called Mixture Premium. It can be defined as follows.

**Definition 2.** *The function* $MP_g : [0, \infty]^k \to [0, \infty]$ *given by*

$$MP_g(P_1^{max}, P_2^{max}, \cdots, P_k^{max}) = \frac{\sum_{i=1}^{k} g(P_i^{max}) \cdot P_i^{max} \cdot n_i}{\sum_{i=1}^{k} g(P_i^{max}) \cdot n_i}, \tag{18}$$

*where $k$ is the number of alternatives in questionnaire, is called a mixture premium.*

On the basis of the formula (18), we evaluated the aggregated maximum mixture premium $MP_g$, the values of which are in Tab. 3 and Tab. 4, in the last column. In our investigation, we apply weighting function $g : [0, \infty] \to ]0, \infty[$, which is given by $g(x) = x^{-q}$, where $q \geq 0$. Function $g$ is a non-increasing weighting function and the mixture operator (11) is an aggregation operator if $0 \leq q \leq 1$. We have chosen just this weighting function to give the highest weight to the smallest maximum premium obtained from all alternatives. Obviously, the decision maker has also other posibilities to investigate and determine the most suitable maximum premium, for example, to use parameter $q \in ]0, 1[$, or other weighting functions. In our future work, we plan to study aggregation of maximum premiums according to the distances of individual premiums from the expected loss.

## 5 Conclusion

In our paper, we proposed a process of utility calibration on the basis of expected utility maximization. Moreover, we have created individual personal utility functions for the determination of maximum premiums in the case of an insurance of wealth worth 12,000 euros.

In our next study we plan to aggregate the obtained premiums by so-called Density-based Averaging operator [1], [3] where variable weights depend on the relative density of a data sample, in our case on the relative distances of individual premiums with different risk aversion coefficients, and moreover, on the relative distances from the expected loss.

The expected utilities and the correspondig maximum premiums and mixture premiums have been evaluated by MS Office Excel 2010 and Mathematica 8 systems.

**Acknowledgments.** This paper is based upon work supported by the Project: Mobility - Enhancing Research, Science and Education at Matej Bel University, ITMS code: 26110230082, under the Operational Program Education co-financed by the European Social Fund.

## References

1. Angelov, P., Yager, R.: Density-based averaging - A new operator for data fusion. Information Sciences 222, 163–174 (2013)
2. Beliakov, G., Pradera, A., Calvo, T.: Aggregation Functions: A Guide for Practitioners. STUDFUZZ, vol. 221. Springer, Heidelberg (2007)
3. Beliakov, G., Wilkin, T.: On some properties of weighted averaging with variable weights. Submitted to Information Sciences (2013)
4. Brunovský, P.: Microeconomics (in Slovak), `http://www.iam.fmph.uniba.sk/skripta/brunovsky2/PART0.pdf`.
5. Bustince, H., Herrera, F., Montero, J. (eds.): Fuzzy Sets and their Extensions: Representation, Aggregation and Models. STUDFUZZ, vol. 220. Springer, New York (2008)
6. Calvo, T., Mesiar, R., Yager, R.R.: Quantitative Weights and Aggregation. IEEE Transaction on Fuzzy Systems 12(1), 62–69 (2004)
7. Grabisch, M., Marichal, J.L., Mesiar, R., Pap, E.: Aggregation functions. Encyclopedia of Mathematics ad Its Foundations. Cambridge University Press (2009)
8. Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M.: Modern Actuarial Risk Theory. Kluwer Academic Publishers, Boston (2001)
9. Khurshid, A.: The end of rationality?, Abstracts. In: Tenth International Conference on Fuzzy Set Theory and Applications, FSTA 2010, Slovak Republic, pp. 10–11 (2010)
10. Melicherčík, I.: Some useful optimization problems in portfolio theory, `http://www.defm.fmph.uniba.sk/ludia/melichercik/papers/jarna_skola_2012.pdf`
11. Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, US (1947)

12. Neumann, J., et al.: Theory of Games and Economic Behavior, 60th Anniversary edn. Princeton University Press, US (2007)
13. Rotar, I.V.: Actuarial Models: The Mathematics of Insurance. Chapman & Hall/CRC Press, Boca Raton (2007)
14. Špirková, J.: Dissertation thesis, Weighted aggregation operators and their applications, Slovak University of Technology in Bratislava, Faculty of Civil Engineering (2008)
15. Špirková, J.: Mixture utility in General Insurance. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) IPMU 2010. CCIS, vol. 80, pp. 712–721. Springer, Heidelberg (2010)
16. Urbaníková, M., Vaculíková, Ľ.: Actuarial mathematics, Slovak University of Technology, Bratislava (2006) (in Slovak)
17. Yager, R.R.: Generalized OWA Aggregation Operators. Fuzzy Optimization and Decision Making 3, 93–107 (2004)

# The General Expression of the Prior Convergence Error: A Proof

Janneke H. Bolt

Department of Information and Computing Sciences, Utrecht University
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
janneke@cs.uu.nl

**Abstract.** In [2], we introduced the notion of the *parental synergy*. In the same paper, moreover, an expression was advanced for the prior convergence error (the error which is found in the marginal probabilities computed for a node when the parents of this node are wrongfully assumed to be independent), in which the parental synergy has a key position as weighting factor. This key position suggests that the parental synergy captures a fundamental feature of a Bayesian network. In this paper a proof is provided for the correctness of the conjectured expression of the prior convergence error.

## 1 Introduction

A Bayesian network is a concise representation of a joint probability distribution over a set of stochastic variables, consisting of a directed acyclic graph and a set of conditional probability distributions [4]. The nodes of the graph represent the variables of the distribution and the arcs of the graph capture (conditional) independencies. From a Bayesian network, in theory, any probability of the represented distribution can be inferred. Since computation of the correct probabilities requires that the dependencies between the variables are taken into account, inference is NP-hard in general [3]. The design of a Bayesian network is such that in the computation of the marginal probabilities of a node just the dependencies between its parent nodes have to be considered. In [1] we termed the error which may arise in the computed marginals when these dependencies are neglected the prior convergence error.

In [2] we introduced the notion of the parental synergy. The parental synergies of a node are computed from the parameters as specified for this node in a Bayesian network. In the same paper, we conjectured an expression for the prior convergence error for the general case of a child node with an arbitrary number of dependent parent nodes. The proposed expression is of interest because of its structure. It includes a part that captures the degree of the dependency between the parent nodes, and a part composed of the parental synergies of the node. In the expression of the prior convergence error these parental synergies act as weighting factors, determining to what extent the degree of dependency between the parent nodes can affect the computed probabilities. We stated that

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 496–505, 2014.

the role of the parental synergy in the expression of the prior convergence error suggests that it captures a fundamental feature of a Bayesian network. In this respect, we note that the parental synergy is related to the concepts of qualitative influence and additive synergy as defined for qualitative probabilistic networks by Wellman [5]. The concept of parental synergy, however, is more general and more informative. In this paper, we provide a proof of the correctness of the expression of the prior convergence error that we conjectured in [2].

## 2 General Preliminaries

We will use the following notation: Variables are denoted by upper-case letters ($A$, $A_k$), and their values by lower-case letters ($a_i, a_{i_k}$); sets of variables by bold-face upper-case letters ($\mathbf{A}, \mathbf{A_k}$) and their instantiations by bold-face lower-case letters ($\mathbf{a_i}, \mathbf{a_{i_k}}$). An arbitrary value assignment to $A$ may also be indicated by $a$ instead of $a_i$ and an arbitrary joint value assignment to $\mathbf{A}$ may also be indicated by $\mathbf{a}$ instead of $\mathbf{a_i}$. The upper-case letter is also used to indicate the whole range of values or value combinations of a variable or a set of variables.

Figure 1 depicts the graph of a Bayesian network. The network includes a node $C$ with $n$, possibly dependent, parents $\mathbf{A} = A_1, \ldots, A_n$, $n \geq 0$. In this network, the marginal probability $\Pr(c)$, of an arbitrary value $c$ of $C$, equals

$$\Pr(c) = \sum_{\mathbf{A}} \Pr(c \mid \mathbf{A}) \cdot \Pr(\mathbf{A})$$

An approximation of this probability can be found by assuming independence of the parents $\mathbf{A}$, which yields the probability

$$\widetilde{\Pr}(c) = \sum_{\mathbf{A}} \Pr(c \mid \mathbf{A}) \cdot \Pr(A_1) \cdot \ldots \cdot \Pr(A_n)$$

In [1], we defined the prior convergence to equal $\Pr(c) - \widetilde{\Pr}(c)$, and in [2], we conjectured an expression for the prior convergence error given a node with an arbitrary number of arbitrary-valued, possibly dependent, parent nodes, as depicted in Figure 1. We restate this expression below.

The parental synergy, a notion that we introduced in [2], is an important factor in the conjectured expression of the prior convergence error and in the definition of the parental synergy an indicator function, called $\delta$, is used. The definitions of this indicator function and of the parental synergy are stated first.

**Definition 1. (The Indicator Function $\delta$)** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be disjoint sets of variables. The indicator function* $\delta$ *on the joint value assignments* $a_{i_1}, \ldots, a_{i_n}$ *to the set of variables* $A = A_1, \ldots, A_n$, $n \geq 0$, *given a specific assignment* $a_{s_1}, \ldots, a_{s_n}$ *and an arbitrary value assignments* $\mathbf{b}$ *is:*

$$\delta(a_{i_1}, \ldots, a_{i_n} \mid a_{s_1}, \ldots, a_{s_n}\mathbf{b}) = \begin{cases} 1 & \text{if } \sum_{k=1,\ldots,n} a_{i_k} \neq a_{s_k} \text{ is even} \\ -1 & \text{if } \sum_{k=1,\ldots,n} a_{i_k} \neq a_{s_k} \text{ is odd} \end{cases}$$

*where* true $\equiv 1$ *and* false $\equiv 0$. □

The indicator function compares the joint value assignment $a_{i_1}, \ldots, a_{i_n}$ with the assignment $a_{s_1}, \ldots, a_{s_n}$, and counts the number of differences: the assignment $a_{i_1}, \ldots, a_{i_n}$ is mapped to the value 1 if the number of differences is even and is mapped to $-1$ if the number of differences is odd. We note that although $\mathbf{b}$ has no influence on the outcome of $\delta$, for notational reasons it is convenient if the function allows for this arbitrary value assignment.

**Definition 2. (The Parental Synergy)** *Let $\mathcal{B}$ be a Bayesian network, representing a joint probability distribution* Pr *over a set of variables $\mathbf{V}$. Let $\mathbf{A} = \{A_1, \ldots, A_n\} \subseteq \mathbf{V}, n \geq 0$, and let $C \in \mathbf{V}$ such that $C$ is a child of all variables in the set $\mathbf{A}$, that is, $A_j \rightarrow C, j = 1, \ldots, n$. Let $\mathbf{a}$ be a joint value assignment to $\mathbf{A}$ and let $c$ be a value of $C$. Furthermore, let $\mathbf{X} \subseteq \rho(C) \backslash \mathbf{A}$, where $\rho(C)$ denotes the parents of $C$, and let $\mathbf{x}$ be a value assignment to $\mathbf{X}$. The* parental synergy *of $\mathbf{a}$ with respect to $c$ given $\mathbf{X} = \mathbf{x}$, denoted as $Y_{\mathbf{x}}(c \mid \mathbf{a})$, is*

$$Y_{\mathbf{x}}(c \mid \mathbf{a}) = \sum_{\mathbf{A}} \delta(\mathbf{a} \mid \mathbf{A}) \cdot \Pr(c \mid \mathbf{A}\mathbf{x})$$

$$Y_{\mathbf{x}}(c) = \Pr(c \mid \mathbf{x}) \qquad \square$$

**Example 1.** *For a node $C$ with parents $A$ with values $a_1$, $a_2$ and $a_3$ and $B$ with values $b_1$ and $b_2$, with $\Pr(c \mid a_1 b_1) = r$, $\Pr(c \mid a_1 b_2) = s$, $\Pr(c \mid a_2 b_1) = t$, $\Pr(c \mid a_2 b_2) = u$, $\Pr(c \mid a_3 b_1) = v$ and $\Pr(c \mid a_3 b_2) = w$, $Y(c \mid a_1 b_1) = r - s - t + u - v + w$, $Y_{a_2}(c \mid b_2) = -t + u$ and $Y_{a_1 b_2}(c) = s$.* $\square$

Note that the parental synergy is related to the concepts of qualitative influence and additive synergy as defined for qualitative probabilistic networks by Wellman [5]. Most obviously, in a binary network, given a node $C$ with a single parent $A$, the sign of the qualitative influence between $A$ and $C$ is derived from $\Pr(c \mid a) - \Pr(c \mid \bar{a})$, which equals $Y(c \mid a)$; given a node $C$ with just the parents $A$ and $B$ the sign of the additive synergy of $A$ and $B$ with respect to $C$ is derived from $\Pr(c \mid ab) - \Pr(c \mid a\bar{b}) - \Pr(c \mid \bar{a}b) + \Pr(c \mid \bar{a}\bar{b})$, which equals $Y(c \mid ab)$. The parental synergy, however is more general since it is defined for an arbitrary number of parent nodes whereas the qualitative influence concerns the interaction between a child node and one parent node, the additive synergy concerns the interactions between a child node and two parent nodes and no analogous concepts are defined for interactions between a child node and more than two parent nodes. Moreover, the parental synergy is more informative, since it yields a number whereas the qualitative influence and the additive synergy are given by a '+', a '−' or the uninformative sign '?'.

In [2], we conjectured an expression for prior convergence error $\Pr(c) - \widetilde{\Pr}(c)$. This expression is stated in the following theorem.

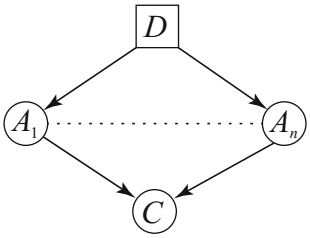



**Fig. 1.** A graph of a Bayesian network with a node $C$ with the, possibly dependent, parents $A_1, \ldots, A_n$

**Theorem 1.** *Let $\mathcal{B}$ be a Bayesian network, representing a joint probability distribution* Pr *over a set of arbitrary-valued variables $\mathbf{V}$. Let $C \in \mathbf{V}$ and let $\rho(C) = \mathbf{A} = \{A_1, \ldots, A_n\} \subseteq \mathbf{V}, n \geq 0$ be the set of, possibly dependent, parents of $C$. The prior convergence error $\Pr(c) - \widetilde{\Pr}(c)$ then equals*

$$\sum_{\mathbf{m}} \Big[ \sum_{\mathbf{A_m}} \Big( (\Pr(A_x \ldots A_y) - \Pr(A_x) \cdot \ldots \cdot \Pr(A_y)) \cdot \sum_{\mathbf{A} \setminus \mathbf{A_m}} Y_{\mathbf{A} \setminus \mathbf{A_m}}(c \mid \mathbf{A_m}) \Big) \Big] / 2^n$$

where

$$\{\mathbf{m}\} = \mathcal{P}(\{1, \ldots, n\})$$
$$\mathbf{A_m} = \{A_x \ldots A_y\}, \mathbf{m} = \{x, \ldots, y\}$$ □

By the summation over $\mathbf{m}$, all combinations of parent nodes are considered, moreover, by the summation over $\mathbf{A_m}$ for each combination of parent nodes all combinations of value assignments to these parent nodes are taken into account. In the remainder of the paper, we will keep using these notations $\mathbf{m}$ and $\mathbf{A_m}$.

The expression above is of theoretical value. It shows that the parental synergy is a weighting factor that determines the impact of the degree of dependency between the parent nodes for a given value assignment, as reflected by $\Pr(A_x \ldots A_y) - \Pr(A_x) \cdot \ldots \cdot \Pr(A_y)$ on the size of the convergence error. Note that if the number of elements of $\mathbf{m}$ is smaller than two, that is, if just one parent or zero parents are considered, then $\Pr(A_x \ldots A_y)$ equals $\Pr(A_x) \cdot \ldots \cdot \Pr(A_y)$ and thus $\Pr(A_x \ldots A_y) - \Pr(A_x) \cdot \ldots \cdot \Pr(A_y)$ equals zero.

## 3 The Prior Convergence Error Illustrated

Given a node with just two, binary, parent nodes, the prior convergence error and the role of the parental synergy can be illuminated graphically. Consider Figure 1 and consider that $n = 2$, that $A_1$ and $A_2$ are binary. The expression of the prior convergence error then reduces to

$$\begin{aligned}\Pr(c) - \widetilde{\Pr}(c) &= [\sum_{A_1A_2} \big(\Pr(A_1A_2) - \Pr(A_1)\cdot\Pr(A_2)\big)\cdot Y(c \mid A_1A_2)]/4 \\ &\quad \{\text{using that } \Pr(a_1a_2) + \Pr(a_1\bar{a}_2) + \Pr(\bar{a}_1a_2) + \Pr(\bar{a}_1\bar{a}_2) = 1 \\ &\quad \text{and that } \Pr(a_i) = \Pr(a_ia_j) + \Pr(a_i\bar{a}_j)\} \\ &= \big(\Pr(a_1a_2) - \Pr(a_1)\cdot\Pr(a_2)\big)\cdot Y(c \mid a_1a_2)\end{aligned}$$

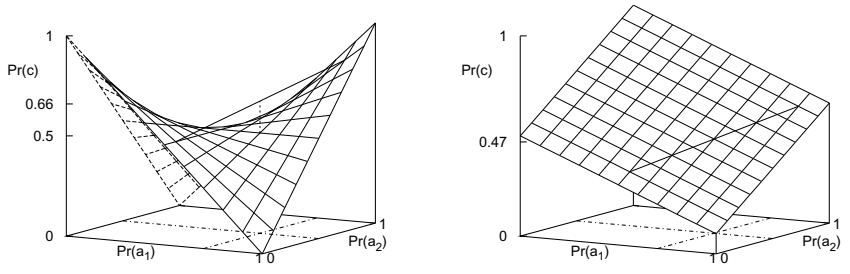
which can, using that $\Pr(a_1a_2) = \Pr(a_1 \mid d)\cdot\Pr(a_2 \mid d)\cdot\Pr(d) + \Pr(a_1 \mid \bar{d})\cdot\Pr(a_2 \mid \bar{d})\cdot\Pr(\bar{d})$ and that $\Pr(a_i) = \Pr(a_i \mid d)\cdot\Pr(d) + \Pr(a_i \mid \bar{d})\cdot\Pr(\bar{d})$, be written as

$$\Pr(c) - \widetilde{\Pr}(c) = l \cdot m \cdot n \cdot Y(c \mid a_1a_2)$$

where

$$\begin{aligned} l &= \Pr(d) - \Pr(d)^2 \\ m &= \Pr(a_1 \mid d) - \Pr(a_1 \mid \bar{d}) \\ n &= \Pr(a_2 \mid d) - \Pr(a_2 \mid \bar{d}) \end{aligned}$$

Now consider for example that $\Pr(d) = 0.5$, $\Pr(a_1 \mid d) = 0.5, \Pr(a_1 \mid \bar{d}) = 0.9, \Pr(a_2 \mid d) = 0.1, \Pr(a_2 \mid \bar{d}) = 0.9, \Pr(c \mid a_1a_2) = 1, \Pr(c \mid a_1\bar{a}_2) = 0, \Pr(c \mid \bar{a}_1a_2) = 0$ and $\Pr(c \mid \bar{a}_1\bar{a}_2) = 1$. The prior convergence error for this example is illustrated in Figure 2a. The line segment in this figure captures the exact probability $\Pr(c)$ as a function of $\Pr(d)$. $\Pr(d)$ itself is not indicated in the figure, note however, that each particular $\Pr(d)$ has a corresponding $\Pr(a_1)$ and $\Pr(a_2)$. The end points of the line segment, for example, are found at $\Pr(d) = 1$ with the corresponding $\Pr(a_1) = 0.5$ and $\Pr(a_2) = 0.1$ and at $\Pr(d) = 0$ with the corresponding $\Pr(a_1) = 0.9$ and $\Pr(a_2) = 0.9$. The surface captures $\widetilde{\Pr}(c)$ as a function of $\Pr(a_1)$ and $\Pr(a_2)$. The convergence error equals the distance between the point on the line segment that matches the probability $\Pr(d)$ from the network and its orthogonal projection on the surface. For $\Pr(d) = 0.5$ the difference between $\Pr(c)$ and $\widetilde{\Pr}(c)$ is indicated by the vertical dotted line segment and equals $0.66 - 0.5 = 0.16$. The factor $l$ reflects the location of the point with the exact probability on the line segment and the factors $m$ and $n$ reflect the location of the line segment. The parental synergy $Y(c \mid a_1a_2)$ now reflects the curvature of the surface with the approximate probabilities. The curvature of the surface determines to what extent the dependency between $A_1$ and $A_2$ can affect the computed probabilities. In the example, the curvature of the surface is maximal. Figure 2b shows, in contrast, a situation in which the parental synergy equals zero. In this example the specifications for nodes $D$, $A_1$ and $A_2$ remained the same, but the specification for node $C$ has changed to $\Pr(c \mid a_1a_2) = 0.6, \Pr(c \mid a_1\bar{a}_2) = 0.1, \Pr(c \mid \bar{a}_1a_2) = 1.0$ and $\Pr(c \mid \bar{a}_1\bar{a}_2) = 0.5$. Now, the surface is flat; $Y(c \mid a_1a_2) = 0$ and the prior convergence error equals zero.




(a) (b)

**Fig. 2.** The prior convergence error $\Pr(c) - \widetilde{\Pr}(c)$ illustrated, given a child node with just two, binary, parent nodes $A_1$ and $A_2$, given $Y(c \mid ab) = 2$ (a) and given $Y(c \mid ab) = 0$ (b)

## 4 A Proof the Expression of the Prior Convergence Error

In order to proof the validity of Theorem 1 we propose the following expressions for $\Pr(c)$ and $\widetilde{\Pr}(c)$.

**Proposition 1.** *Let $\mathcal{B}$ be a Bayesian network, representing a joint probability distribution* Pr *over a set of variables $\mathbf{V}$. Let $C \in \mathbf{V}$ and let $\rho(C) = \mathbf{A} = \{A_1, \ldots, A_n\} \subseteq \mathbf{V}, n \geq 0$ be the set of, possibly dependent, parents of $C$. The prior probability* $\Pr(c)$ *then equals*

$$\Pr(c) = \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \Pr(\mathbf{A_m}) \cdot \sum_{\mathbf{A} \setminus \mathbf{A_m}} Y_{\mathbf{A} \setminus \mathbf{A_m}}(c \mid \mathbf{A_m})/2^n \tag{1}$$

**Example 2.** For a variable $C$ with the parents $A_1$ and $A_2$, according to Proposition 1

$$4 \cdot \Pr(c) = \sum_{A_1 A_2} \Pr(A_1 A_2) \cdot Y(c \mid A_1 A_2) + \sum_{A_1} \Pr(A_1) \cdot \sum_{A_2} Y_{A_2}(c \mid A_1) + \sum_{A_2} \Pr(A_2) \cdot \sum_{A_1} Y_{A_1}(c \mid A_2) + \sum_{A_1 A_2} Y_{A_1 A_2}(c)$$

□

Given Proposition 1, the approximation $\widetilde{\Pr}(c)$ can be written as:

$$\widetilde{\Pr}(c) = \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \Pr(A_x) \cdot \ldots \cdot \Pr(A_y) \cdot \sum_{\mathbf{A} \setminus \mathbf{A_m}} Y_{\mathbf{A} \setminus \mathbf{A_m}}(c \mid \mathbf{A_m})/2^n \tag{2}$$

In the proof of Proposition 1 we will use the following lemma:

**Lemma 1.** *Let* $\mathbf{A} = \{A_1, \cdots, A_n\}$ *be a set of variables and let* $\mathbf{a} = a_1, \ldots, a_n$ *be an arbitrary given joint value assignment to* $\mathbf{A}$*. Then* $2^n \cdot \Pr(\mathbf{a})$ *equals*

$$2^n \cdot \Pr(\mathbf{a}) = \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \Pr(\mathbf{A_m}) \cdot \delta(\mathbf{A_m} \mid \mathbf{a}) \qquad \square$$

**Example 3.** Given the variables $A$ with the values $a$ and $\bar{a}$ and $B$ with the values $b$ and $\bar{b}$, according to Lemma 1, $4 \cdot \Pr(ab) = \Pr(ab) - \Pr(a\bar{b}) - \Pr(\bar{a}b) + \Pr(\bar{a}\bar{b}) + \Pr(a) - \Pr(\bar{a}) + \Pr(b) - \Pr(\bar{b}) + 1$ $\square$

**Proof of Lemma 1.** We first rewrite

$$\begin{aligned} 2^n \cdot \Pr(\mathbf{a}) &= \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \Pr(\mathbf{A_m}) \cdot \delta(\mathbf{A_m} \mid \mathbf{a}) \\ &= \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \sum_{\mathbf{A} \backslash \mathbf{A_m}} \Pr(\mathbf{A} \backslash \mathbf{A_m}, \mathbf{A_m}) \cdot \delta(\mathbf{A_m} \mid \mathbf{a}) \\ &= \sum_{\mathbf{m}} \sum_{\mathbf{A}} \Pr(\mathbf{A}) \cdot \delta(\mathbf{A_m} \mid \mathbf{a}) \end{aligned}$$

In the expression $\sum_{\mathbf{m}} \sum_{\mathbf{A}} \Pr(\mathbf{A}) \cdot \delta(\mathbf{A_m} \mid \mathbf{a})$ we have that $\sum_{\mathbf{m}}$ selects all possible combinations of the variables of $\mathbf{A}$, and $\sum_{\mathbf{A}}$ sums, for each of those combinations, $\Pr(\mathbf{A}) \cdot \delta(\mathbf{A_m} \mid \mathbf{a})$. The outcome of the function $\delta$ is determined by the variables selected by $\mathbf{m}$ with the values as determined by $\sum_{\mathbf{A}}$, and the values of $\mathbf{A}$ in the given joint value assignment $\mathbf{a}$.

**Example 4.** Given the variables $A$ and $B$ and the value assignment $A = a$ and $B = b$, the expression

$$\sum_{\mathbf{m}} \sum_{\mathbf{A}} \Pr(\mathbf{A}) \cdot \delta(\mathbf{A_m} \mid \mathbf{a})$$

results in:

$$\sum_{AB} \Pr(AB) \cdot \delta(AB \mid ab) + \sum_{AB} \Pr(AB) \cdot \delta(A \mid ab) + \sum_{AB} \Pr(AB) \cdot \delta(B \mid ab) + \sum_{AB} \Pr(AB) \cdot \delta(. \mid ab)$$

$\square$

Now divide $\{\mathbf{m}\}$ in $\{\mathbf{m_{-1}}\} = \{\mathbf{m} \mid 1 \notin \mathbf{m}\}$ and $\{\mathbf{m_{+1}}\} = \{\mathbf{m} \mid 1 \in \mathbf{m}\}$. Thus, $\mathbf{m_{-1}}$ selects all possible subsets from $\mathbf{A}$ in which $A_1$ is included and $\mathbf{m_{+1}}$ selects all possible subsets from $\mathbf{A}$ without $A_1$. These two sets of subsets of $\mathbf{A}$ include, in pairs, the same subsets, apart from $A_1$. Thus, $\emptyset$ is selected by $\mathbf{m_{-1}}$

and $\{A_1\}$ is selected by $\mathbf{m_{+1}}$; $\{A_2\}$ is selected by $\mathbf{m_{-1}}$ and $\{A_1, A_2\}$ is selected by $\mathbf{m_{+1}}$, etcetera. For each of those pairs, we find for all value combinations $\mathbf{A}$ with $A_1 \neq a_1$ that

$$\delta(\mathbf{A_{m_{-1}}} \mid \mathbf{a}) = -\delta(\mathbf{A_{m_{+1}}} \mid \mathbf{a})$$

(Remember that $\mathbf{a} = a_1, \ldots, a_n$. The outcome of $\delta(\mathbf{A_{m_{-1}}} \mid \mathbf{a})$ is, apart from $A_1$, determined by the same value assignments to the same variables as the outcome of $\delta(\mathbf{A_{m_{+1}}} \mid \mathbf{a})$. Now when $A_1 \neq a_1$ there is one extra difference in value assignment, which changes the sign of the outcome of $\delta$.) Thus under the condition that $A_1 \neq a_1$ we find that

$$\sum_{\mathbf{m_{+1}}} \sum_{\mathbf{A}} \Pr(\mathbf{A}) \cdot \delta(\mathbf{A_{m_{+1}}} \mid \mathbf{a}) + \sum_{\mathbf{m_{-1}}} \sum_{\mathbf{A}} \Pr(\mathbf{A}) \cdot \delta(\mathbf{A_{m_{-1}}} \mid \mathbf{a}) = 0$$

This implies that we only have to consider $A_1 = a_1$ and thus that

$$\begin{aligned} 2^n \cdot \Pr(\mathbf{a}) &= \sum_{\mathbf{m}} \sum_{\mathbf{A}} \Pr(\mathbf{A}) \cdot \delta(\mathbf{A_m} \mid \mathbf{a}) \\ &= \sum_{\mathbf{m_{+1}}} \sum_{\mathbf{A}} \Pr(\mathbf{A}) \cdot \delta(\mathbf{A_{m_{+1}}} \mid \mathbf{a}) + \sum_{\mathbf{m_{-1}}} \sum_{\mathbf{A}} \Pr(\mathbf{A}) \cdot \delta(\mathbf{A_{m_{-1}}} \mid \mathbf{a}) \\ &= \sum_{\mathbf{m}} \sum_{\mathbf{A}/A_1} \Pr(\mathbf{A}/A_1, a_1) \cdot \delta(\mathbf{A_m}(/A_1, a_1) \mid \mathbf{a}) \end{aligned}$$

where

$$\delta(\mathbf{A_m}(/A_1, a_1) \mid \mathbf{a}) = \begin{cases} \delta(\mathbf{A_m}/A_1, a_1 \mid \mathbf{a}) & \text{if } 1 \in \mathbf{m} \\ \delta(\mathbf{A_m} \mid \mathbf{a}) & \text{if } 1 \notin \mathbf{m} \end{cases}$$

In a next step we divide $\mathbf{m}$ in $\mathbf{m_{-2}}$ and $\mathbf{m}$ in $\mathbf{m_{+2}}$. We have, analogous to the division of $\mathbf{m}$ in $\mathbf{m_{-1}}$ and $\mathbf{m}$ in $\mathbf{m_{+1}}$, that for all value combinations $\mathbf{A}/A_1$, given that $A_2 \neq a_2$

$$\delta(\mathbf{A_{m_{-2}}}(/A_1, a_1) \mid \mathbf{a}) = -\delta(\mathbf{A_{m_{+2}}}(/A_1, a_1) \mid \mathbf{a})$$

And thus we have that:

$$\begin{aligned} 2^n \cdot \Pr(\mathbf{a}) &= \sum_{\mathbf{m}} \sum_{\mathbf{A}/A_1} \Pr(\mathbf{A}/A_1, a_1) \cdot \delta(\mathbf{A_m}(/A_1, a_1) \mid \mathbf{a}) \\ &= \sum_{\mathbf{m}} \sum_{\mathbf{A}/A_1 A_2} \Pr(\mathbf{A}/A_1 A_2, a_1 a_2) \cdot \delta(\mathbf{A_m}(/A_1 A_2, a_1 a_2) \mid \mathbf{a}) \end{aligned}$$

This pinning down of the value of a variable in the summation over $\mathbf{A}$ can be repeated for all $n$ which implies that

$$\begin{aligned} 2^n \cdot \Pr(\mathbf{a}) &= \sum_{\mathbf{m}} \sum_{\mathbf{A}} \Pr(\mathbf{A}) \cdot \delta(\mathbf{A_m} \mid \mathbf{a}) \\ &= \sum_{\mathbf{m}} \Pr(\mathbf{a}) \cdot \delta(\mathbf{a_m} \mid \mathbf{a}) \end{aligned}$$

which concludes the proof of Lemma 1. □

**Proof of Proposition 1.** In order to prove Proposition 1 we multiply left and right hand of Expression 1 with $2^n$ and rewrite the resulting right hand using the definition of the parental synergy.

$$\begin{aligned} 2^n \cdot \Pr(c) &= \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \Pr(\mathbf{A_m}) \cdot \sum_{\mathbf{A \backslash A_m}} Y_{\mathbf{A \backslash A_m}}(c \mid \mathbf{A_m}) \\ &= \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \Pr(\mathbf{A_m}) \cdot \sum_{\mathbf{A \backslash A_m}} \sum_{\mathbf{A^*_m}} \delta(\mathbf{A_m} \mid \mathbf{A^*_m}) \cdot \Pr(c \mid \mathbf{A^*_m}, \mathbf{A \backslash A_m}) \\ &= \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \Pr(\mathbf{A_m}) \cdot \sum_{\mathbf{A}} \delta(\mathbf{A_m} \mid \mathbf{A}) \cdot \Pr(c \mid \mathbf{A}) \end{aligned}$$

In the second step above, an asterisk is used in order to distinguish between the two different summations over $\mathbf{A_m}$. In the next step, $\mathbf{A^*_m}$ and $\mathbf{A \backslash A_m}$ are combined to $\mathbf{A}$ after which an asterisk is not needed any more to indicate the distinction. Note that (for notational reasons), $\delta(\mathbf{A_m} \mid \mathbf{A^*_m})$ is changed to $\delta(\mathbf{A_m} \mid \mathbf{A})$, which has the same outcome. Simply rearranging terms and dividing the left and the right hand of the equation by $2^n$ now results in the following form of Proposition 1

$$\Pr(c) = \sum_{\mathbf{A}} \Pr(c \mid \mathbf{A}) \cdot \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \Pr(\mathbf{A_m}) \cdot \delta(\mathbf{A_m} \mid \mathbf{A}) / 2^n$$

Since, by definition

$$\Pr(c) = \sum_{\mathbf{A}} \Pr(c \mid \mathbf{A}) \cdot \Pr(\mathbf{A})$$

and, by Lemma 1,

$$2^n \cdot \Pr(\mathbf{A}) = \sum_{\mathbf{m}} \sum_{\mathbf{A_m}} \Pr(\mathbf{A_m}) \cdot \delta(\mathbf{A_m} \mid \mathbf{A})$$

rewriting Proposition 1, together with the proof of Lemma 1, provides the proof of Proposition 1. □

**Proof of Theorem 1.** The validity of Theorem 1 follows from the validity of Proposition 1 and its consequence Equation 2. □

## 5 Discussion

In [2], we conjectured an expression for the prior convergence error. The prior convergence error is the error which is found in the marginal prior probabilities computed for a node in a Bayesian network when the parents of this node are wrongfully assumed to be independent. The proposed expression is interesting because of its structure. The expression consists of a part that captures the degree of dependency between the parents of the node, and furthermore it includes the parental synergies of the node. The parental synergies are computed from the conditional probabilities as specified for the node in a Bayesian network and act as a weighting factor, determining to what extent the degree of dependency between the parent nodes can affect the computed probabilities. The role of the parental synergy in the expression of the prior convergence error suggests that it captures a fundamental feature of a Bayesian network. In this respect, we noted that the parental synergy is related to the concepts of qualitative influence and additive synergy as defined for qualitative probabilistic networks by Wellman but is more general and more informative. In this paper we provided a proof of the correctness of the expression of the prior convergence error that we conjectured in [2].

**Acknowledgments.** This research was supported by the Netherlands Organisation for Scientific Research (NWO).

## References

1. Bolt, J.H., van der Gaag, L.C.: The convergence error in loopy propagation. Paper presented at the International Conference on Advances in Intelligent Systems: Theory and Applications (2004)
2. Bolt, J.H.: Bayesian networks: the parental synergy. In: Jaeger, M., Nielsen, T.D. (eds.) Proceedings of the Fourth European Workshop on Probabilistic Graphical Models, Hirtshals, pp. 33–40 (2008)
3. Cooper, G.F.: The computational complexity of probabilistic inference using Bayesian belief networks. Artificial Intelligence 42, 393–405 (1990)
4. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, Palo Alto (1988)
5. Wellman, M.P.: Fundamental concepts of qualitative probabilistic networks. Artificial Intelligence 44, 257–303 (1990)

# On SPI for Evaluating Influence Diagrams

Rafael Cabañas[1], Anders L. Madsen[2], Andrés Cano[1], and Manuel Gómez-Olmedo[1]

[1] Department of Computer Science and Artificial Intelligence CITIC,
University of Granada, Spain
{rcabanas,mgomez,acu}@decsai.ugr.es
[2] HUGIN EXPERT A/S and Department of Computer Science,
Aalborg University Aalborg, Denmark
anders@hugin.com

**Abstract.** An Influence Diagram is a probabilistic graphical model used to represent and solve decision problems under uncertainty. Its evaluation requires to perform a series of combinations and marginalizations with the potentials attached to the Influence Diagram. Finding an optimal order for these operations, which is NP-hard, is an element of crucial importance for the efficiency of the evaluation. The SPI algorithm considers the evaluation as a combinatorial factorization problem. In this paper, we describe how the principles of SPI can be used to solve Influence Diagrams. We also include an evaluation of different combination selection heuristics and a comparison with the variable elimination algorithm.

**Keywords:** Influence Diagrams, Combinatorial Factorization Problem, Exact Evaluation, Heuristic Algorithm.

## 1 Introduction

Influence Diagrams (IDs) [1,2] provide a framework to model decision problems with uncertainty for a single decision maker. The goal of evaluating an ID is to obtain the best option for the decision maker (*optimal policy*) and its utility.

Most of the evaluation algorithms proposed in the literature [3,4,5,6] require to perform a series of combinations and marginalizations with the probability and utility functions (potentials). Finding an optimal order for these operations, which is NP-hard [7], is an element of crucial importance for the efficiency of the evaluation. Thus the evaluation of an ID can be seen as a combinatorial factorization problem. This idea was already used to make inference in Bayesian Networks (BNs) with the first version of Symbolic Probabilistic Inference algorithm (SPI) [8] and with an improved algorithm in the SPI family called set-factoring [9]. In a related work [10] some experiments with SPI were performed to evaluate decision networks, however it was not given any detail of the algorithm. In this paper we describe the SPI algorithm for evaluating IDs taking into account the differences of an ID: two kind of potentials, the temporal order between decisions, etc. The experimental work shows how SPI can improve the efficiency of the evaluation on some IDs and different combination selection heuristics are compared.

The paper is organized as follows: Section 2 introduces some basic concepts about IDs and the motivation of this work; Section 3 describes how SPI can be used for evaluating IDs; Section 4 includes the experimental work and results; finally Section 5 details our conclusions and lines for future work.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 506–516, 2014.

## 2 Preliminaries

### 2.1 Influence Diagrams

An ID [1,2] is a probabilistic graphical model for decision analysis under uncertainty with three kinds of nodes: *chance nodes* (circles) that represent random variables; *decision nodes* (squares) that correspond with the actions which the decision maker can control; and *utility nodes* (hexagons) that represent decision maker preferences.

We denote by $\mathscr{U}_C$ the set of chance nodes, by $\mathscr{U}_D$ the set of decision nodes, and by $\mathscr{U}_V$ the set of utility nodes. The decision nodes have a temporal order, $D_1,\ldots,D_n$, and the chance nodes are partitioned into a collection of disjoint sets according to when they are observed: $\mathscr{I}_0$ is the set of chance nodes observed before $D_1$, and $\mathscr{I}_i$ is the set of chance nodes observed after decision $D_i$ is taken and before decision $D_{i+1}$ is taken. Finally, $\mathscr{I}_n$ is the set of chance nodes observed after $D_n$. That is, there is a partial order: $\mathscr{I}_0 \prec D_1 \prec \mathscr{I}_1 \prec \cdots \prec D_n \prec \mathscr{I}_n$. Fig. 1 shows an example of an ID.

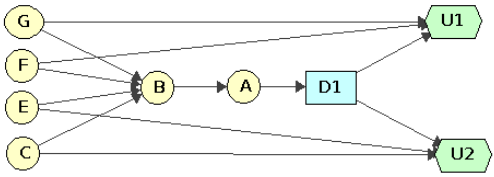



**Fig. 1.** An ID for a decision problem with one decision $D1$. The set of chance variables is partitioned into the sets: $\mathscr{I}_0 = \{A\}$ and $\mathscr{I}_1 = \{B,C,E,F,G\}$. The utility function is a sum of two local utilities, one associated to $D1$, $G$, and $F$ and the other associated to $D1$, $C$, and $E$.

In the description of an ID, it is more convenient to think in terms of predecessors: the parents of a chance node $X_i$, denoted $pa(X_i)$, are also called *conditional predecessors*. The parents of a utility node $V_i$, denoted $pa(V_i)$, are also called conditional predecessors. Similarly, the parents of a decision $D_i$ are called *informational predecessors* and are denoted $pa(D_i)$. Informational predecessors of each decision $D_i$, must include previous decisions and their informational predecessors (*no-forgetting assumption*).

The *universe* of the ID is $\mathscr{U} = \mathscr{U}_C \cup \mathscr{U}_D = \{X_1,\ldots,X_m\}$. Let us suppose that each variable $X_i$ takes values on a finite set $\Omega_{X_i} = \{x_1,\ldots,x_{|\Omega_{X_i}|}\}$. Each chance node $X_i$ has a conditional probability distribution $P(X_i|pa(X_i))$ associated. In the same way, each utility node $V_i$ has a utility function $U(pa(V_i))$ associated. In general, we will talk about potentials (not necessarily normalized). The set of all variables involved in a potential $\phi$ is denoted $dom(\phi)$, defined on $\Omega_{dom(\phi)} = \times\{\Omega_{X_i}|X_i \in dom(\phi)\}$. The elements of $\Omega_{dom(\phi)}$ are called configurations of $\phi$. Therefore, a *probability potential* denoted by $\phi$ is a mapping $\phi : \Omega_{dom(\phi)} \to [0,1]$. A *utility potential* denoted by $\psi$ is a mapping $\psi : \Omega_{dom(\psi)} \to \mathbb{R}$. The set of probability potentials is denoted by $\Phi$ while the set of utility potentials is denoted by $\Psi$.

### 2.2 Motivation

The goal of evaluating an ID is to obtain an *optimal policy* $\delta_i$ for each decision $D_i$, that is a function of a subset of its informational predecessors. The optimal policy maximizes the *expected utility* for the decision.

***Optimal policy:*** *Let ID be an influence diagram over the universe $\mathscr{U} = \mathscr{U}_C \cup \mathscr{U}_D$ and let $\mathscr{U}_V$ be the set of utility nodes. Let the temporal order of the variables be described as $\mathscr{I}_0 \prec D_1 \prec \mathscr{I}_1 \prec \cdots \prec D_n \prec \mathscr{I}_n$. Then, an optimal policy for $D_i$ is*

$$\delta_{D_i}(\mathscr{I}_0, D_1, \ldots, \mathscr{I}_{i-1}) = \arg\max_{D_i} \sum_{\mathscr{I}_i} \max_{D_{i+1}} \cdots \max_{D_n} \sum_{\mathscr{I}_n} \prod_{X \in \mathscr{U}_C} P(X|pa(X)) \left( \sum_{V \in \mathscr{U}_V} U(pa(V)) \right) \quad (1)$$

For example consider the ID shown in Fig. 1. The optimal policy for $D_1$ can be calculated directly from Eq.1:

$$\delta_{D_1}(A) = \arg\max_{D_1} \sum_{G,F,E,C,B} P(G)P(F)P(E)P(C)P(B|C,E,F,G)P(A|B)U \quad (2)$$

where $U = (U_1(G,F,D_1) + U_2(E,C,D_1))$ is the sum of the local utilities. The table representing the joint probability of all chance variables might be too large. For that reason, some evaluation algorithms such as *Variable Elimination* (VE) for IDs [11] re-order the marginalizations of the variables as follows:

$$\delta_{D_1}(A) = \arg\max_{D_1} \sum_{G} P(G) \sum_{F} P(F) \sum_{E} P(E) \sum_{C} P(C) \sum_{B} P(B|C,E,F,G)P(A|B)U \quad (3)$$

Assuming that all the variables are binary and considering only the computations between probability potentials, the calculation of $\delta_{D_1}(A)$ requires 124 multiplications and 64 additions to marginalize out the variables in $\mathscr{I}_0$. Independently of the elimination ordering used to solve this ID, VE will always have to combine the marginal potentials with a large potential such as $P(B|C,E,F,G)$ or $P(A|C,E,F,G)$. However, with a re-order of the operations this situation can be avoided:

$$\delta_{D_1}(A) = \arg\max_{D_1} \sum_{B,C,E} \left( (P(A|B)\,(P(E)P(C))) \left( \sum_{F,G} P(B|C,E,F,G)\,(P(F)P(G))\,U \right) \right) \quad (4)$$

Using Eq.4 the calculation of the optimal policy requires 72 multiplications and 32 additions. In some cases it could be better to combine small potentials even if they do not share any variable (e.g., $P(E)$ and $P(C)$). This combination will never be performed using VE since it is guided by the elimination ordering. Thus the efficiency of the evaluation can be improved if an optimization in the order of both operations, marginalization and combination, is performed [9].

## 3 Symbolic Probabilistic Inference

### 3.1 Overview

As VE does, SPI removes all variables in the decision problem in reverse order of the partial ordering imposed by the information constraints. That is, it first sum-marginalizes

$\mathscr{I}_n$, then max-marginalizes $D_n$, sum-marginalizes $\mathscr{I}_{k-1}$, etc. This type of elimination order is called a strong elimination order [12]. The general scheme of SPI algorithm as presented in this paper is shown in Definition 1.

**Definition 1 (SPI Algorithm)**
*Let ID be an influence diagram. Let $\Phi$ and $\Psi$ be the set of all probability and utility potentials respectively. Let the temporal order of the variables be described as $\mathscr{I}_0 \prec D_1 \prec \mathscr{I}_1 \prec \cdots \prec D_n \prec \mathscr{I}_n$. Then, the procedure for evaluating the ID using SPI algorithm is:*

1. *for $(k := n; k > 0; k := k-1)$*
   (a) *Remove chance variables in $\mathscr{I}_k$ (Definition 2)*
   (b) *Remove decision $D_k$ (Definition 4)*
2. *Remove chance variables in $\mathscr{I}_0$ (Definition 2)*

VE considers the evaluation as a problem of finding the optimal elimination ordering whereas SPI considers it as a combinatorial factorization problem. That is, VE chooses at each step the best variable to remove while SPI chooses the best pair of potentials to combine and eliminate the variables when possible. In this sense SPI is finer grained than VE.

### 3.2 Removal of Chance Variables

In order to remove a subset of chance variables $\mathscr{I}_k$, our version of SPI considers probability and utility potentials separately: first, SPI tries to find the best order for combining all the relevant probability potentials in $\Phi^k$ (potentials containing any of the variables in $\mathscr{I}_k$). For that purpose, all possible pairwise combinations between probability potentials are stored in the set $B$. At each iteration, a pair of probability potentials is selected to be combined. The procedure stops when all variables has been removed. A variable can be removed in the moment it only appears in a single probability potential. This procedure is shown in Definition 2.

**Definition 2 (Removal of a Subset of Chance Variables).** *Let $\mathscr{I}_k$ be the set of variables to remove, let $\Phi$ and $\Psi$ be the set of all current probability and utility potentials respectively in an ID. Then, the procedure for removing $\mathscr{I}_k$ is:*

1. *Set the relevant potentials:*

$$\Phi^k := \{\phi \in \Phi | \mathscr{I}_k \cap dom(\phi) \neq \emptyset\} \quad \Psi^k := \{\psi \in \Psi | \mathscr{I}_k \cap dom(\psi) \neq \emptyset\}$$

2. *Update $\Phi := \Phi \backslash \Phi^k$ and $\Psi := \Psi \backslash \Psi^k$*
3. *Initialize the combination candidate set $B := \emptyset$.*
4. *Repeat:*
   (a) *if $|\Phi^k| > 1$, then*
      i. *Add all pairwise combinations of elements of $\Phi^k$ to $B$ which are not already in $B$.*
      ii. *Select a pair $p := \{\phi_i, \phi_j\}$ of $B$ according to some criteria and combine both potentials: Set $\phi_{ij} := \phi_i \otimes \phi_j$*

*iii. Determine the set* $\mathbf{W}$ *of variables that can be sum-marginalized:*

$$\mathbf{W} := \{X \in dom(\phi_{ij}) \cap \mathscr{I}_k | \forall \phi \in \Phi^k \backslash p : X \notin dom(\phi)\}$$

*iv. Update B by deleting all pairs p where* $\phi_i \in p$ *or* $\phi_j \in p$.
*v. Delete* $\phi_i$ *and* $\phi_j$ *from* $\Phi^k$.
*else*
*i. Let* $\phi_{ij}$ *be the single potential in* $\Phi^k$.
*ii. Determine the set* $\mathbf{W}$ *of variables that can be sum-marginalized:*

$$\mathbf{W} := \{X \in dom(\phi_{ij}) \cap \mathscr{I}_k\}$$

*iii. Delete* $\phi_{ij}$ *from* $\Phi^k$.
*(b) Select the utility potentials relevant for removing* $\mathbf{W}$*:*

$$\Psi^{\mathbf{W}} := \{\psi \in \Psi^k | \mathbf{W} \cap dom(\psi) \neq \emptyset\}$$

*(c) Sum-marginalize variables in* $\mathbf{W}$ *from* $\phi_{ij}$ *and* $\Psi^{\mathbf{W}}$*. A probability potential* $\phi_{ij}^{\downarrow \mathbf{W}}$ *and a set of utility potentials* $\Psi^{\downarrow \mathbf{W}}$ *are obtained as a result (Definition 3).*
*(d) Update the set of variables to remove:* $\mathscr{I}_k := \mathscr{I}_k \backslash \mathbf{W}$
*(e) Update the set of relevant potentials:*

$$\Phi^k := \Phi^k \cup \{\phi_{ij}^{\downarrow \mathbf{W}}\} \qquad \Psi^k := (\Psi^k \setminus \Psi^{\mathbf{W}}) \cup \Psi^{\downarrow \mathbf{W}}$$

*Until* $\mathscr{I}_k = \emptyset$
5. *Update* $\Phi := \Phi \cup \Phi^k$ *and* $\Psi := \Psi \cup \Psi^k$

In Definition 2 only probability potentials are combined while utility potentials are not. Let us suppose that we aim to remove a variable $X$ from a set of probability potentials $\{\phi_1, \ldots, \phi_k, \phi_{ij}\}$ and from a set of utility potentials $\{\psi_1, \ldots, \psi_l, \psi_m, \ldots + \psi_n\}$. Let $\phi_{ij}$ and $\{\psi_m, \ldots, \psi_n\}$ be the potentials containing $X$. Then, the removal of $X$ can be made using Eq.5.

$$\sum_X \phi_1 \cdots \phi_k \phi_{ij} (\psi_1 + \cdots + \psi_l + \psi_m + \cdots + \psi_n) =$$
$$= \phi_1 \cdots \phi_k \left( \sum_X \phi_{ij} \right) \left( \psi_1 + \cdots + \psi_l + \frac{\sum_X (\phi_{ij} (\psi_m + \cdots + \psi_n))}{\sum_X \phi_{ij}} \right) \quad (5)$$

The utility potentials must be combined with $\phi_{ij}$ which is the resulting potential of combining all potentials containing $X$. For that reason, the utilities can only be combined when a variable can be removed. That is the moment when $\phi_{ij}$ has been calculated. The procedure for sum-marginalizing a set of variables (Definition 3) involves finding good order for summing the utility potentials. The procedure for that is quite similar to the procedure for combining probabilities, however the combination candidate set B can contain singletons as well. The reason for that is that in some cases it could be better to apply the distributive law [5,11].

**Definition 3 (Sum-Marginalization).** *Let $\phi$ be a probability potential and $\Psi^{\mathbf{W}}$ a set of utility potentials relevant for removing the chance variables in $\mathbf{W}$. Then, the procedure for sum-marginalizating $\mathbf{W}$ from $\phi$ and $\Psi^{\mathbf{W}}$ is:*

1. *Initialize the combination candidate set $B := \emptyset$.*
2. *Repeat:*
   (a) *Add all pairwise combinations of elements of $\Psi^{\mathbf{W}}$ to B which are not already in B.*
   (b) *Add to B all potentials in $\Psi^{\mathbf{W}}$ that contains any variable of $\mathbf{W}$ which is not present in any other potential of $\Psi^{\mathbf{W}}$, that is a variable that can be removed.*
   (c) *Select a pair $q := \{\psi_i, \psi_j\}$ or a singleton $q := \{\psi_i\}$ from B according to some criteria.*
   (d) *If q is a pair, then $\psi_{ij} := \psi_i + \psi_j$. Otherwise, $\psi_{ij} := \psi_i$*
   (e) *Determine the set $\mathbf{V}$ of variables that can be sum-marginalized:*

$$\mathbf{V} := \{X \in dom(\psi_{ij}) \cap \mathbf{W} | \forall \psi \in \Psi^{\mathbf{W}} \backslash q : X \notin dom(\psi)\}$$

   (f) *Sum-marginalize $\mathbf{V}$, giving as a result:*

$$\phi^{\downarrow \mathbf{V}} := \sum_{\mathbf{V}} \phi \qquad \psi^{\downarrow \mathbf{V}} := \sum_{\mathbf{V}} (\phi \otimes \psi_{ij}) / \phi^{\downarrow \mathbf{V}}$$

   (g) *If q is a pair, remove $\psi_i$ and $\psi_j$ from $\Psi^{\mathbf{W}}$ and any element in B containing them. Otherwise, only remove $\psi_i$ from $\Psi^{\mathbf{W}}$ and any element in B containing it.*
   (h) *Update $\phi := \phi^{\downarrow \mathbf{V}}$ and $\Psi^{\mathbf{W}} := \Psi^{\mathbf{W}} \cup \{\psi^{\downarrow \mathbf{V}}\}$*
   (i) *Update the set of variables to remove: $\mathbf{W} := \mathbf{W} \backslash \mathbf{V}$*

   *Until $\mathbf{W} = \emptyset$*
3. *Return $\phi$ and $\Psi^{\mathbf{W}}$*

### 3.3 Removal of Decision Variables

Once all variables in $\mathscr{I}_k$ are removed using algorithm in Definition 2, a similar procedure must be performed to remove a decision variable $D_k$ (see Definition 4). However, this removal does not imply the combination of any probability potential since any decision is d-separated from its predecessors [11]. Thus, any probability potential $\phi(D_k, \mathbf{X})$ must be directly transform into $\phi(\mathbf{X})$ if $D_k$ is a decision and $\mathbf{X}$ is a set of chance variables that belong to $\mathscr{I}_i$ with $i < k$. This property is used at step 2 of Definition 4.

**Definition 4 (Removal of a Decision Variable).** *Let $D_k$ be the decision variable to remove, let $\Phi$ and $\Psi$ be the set of all current probability and utility potentials respectively. Then, the procedure for removing $D_k$ is:*

1. *Set the relevant potentials:*

$$\Phi^k := \{\phi \in \Phi | D_k \in dom(\phi)\} \qquad \Psi^k := \{\psi \in \Psi | D_k \in dom(\psi)\}$$

2. *Foreach $\phi \in \Phi^k$, remove $D_k$ by restricting $\phi$ to any of the values of $D_k$. The set of potentials $\Phi^{\downarrow D_k}$ is given as a result.*

3. *Max-marginalize variable $D_k$ from $\Psi^k$ and record the policy for $D_k$. A new potential $\psi^{\downarrow D_k}$ is obtained as a result (Definition 5).*
4. *Update the set of potentials in the ID:*

$$\Phi := (\Phi \backslash \Phi^k) \cup \Phi^{\downarrow D_k} \qquad \Psi := (\Psi \backslash \Psi^k) \cup \{\psi^{\downarrow D_k}\}$$

**Definition 5 (Max-Marginalization).** *Let $\Psi$ be a set of utility potentials and $D$ a decision variable. Then, the procedure for max-marginalizating $D$ from $\Psi$ is:*

1. *Initialize the combination candidate set $B := \emptyset$.*
2. *While $|\Psi| > 1$:*
   (a) *Add all pairwise combinations of elements of $\Psi$ to $B$ which are not already in $B$.*
   (b) *Select a pair $q := \{\psi_i, \psi_j\}$ according to some criteria and sum both potentials giving as a result $\psi_{ij}$.*
   (c) *Update $B$ by deleting all pairs $p$ where $\psi_i \in p$ or $\psi_j \in p$.*
   (d) *Update $\Psi := \Psi \backslash \{\psi_i, \psi_j\} \cup \{\psi_{ij}\}$*
3. *Let $\psi^D$ be the single potential in $\Psi$.*
4. *Max-marginalize $D$, giving as a result: $\psi^{\downarrow D} := \max_D \psi^D$*
5. *Return $\psi^{\downarrow D}$*

### 3.4 Heuristics

During the removal of the chance variables, at each iteration a pair of probability potentials is selected to be combined (Definition 2, step 4.a.ii). For that, some heuristics used with VE can be adapted for selecting a pair. Let $p := \{\phi_i; \phi_j\}$ be a candidate pair to be combined, let $\phi_{ij} = \phi_i \otimes \phi_j$ be the resulting potential of the combination and let $\mathbf{W}$ be the set of variables that can be removed. Then, the heuristics *minimum size* [13], *minimum weight* [12] and *Cano and Moral* [14] are defined as:

$$min\_size(p) = |dom(\phi_i) \cup dom(\phi_j)| = |dom(\phi_{ij})| \tag{6}$$

$$min\_weight(p) = \prod_{X \in dom(\phi_{ij})} |\Omega_X| \tag{7}$$

$$Cano\_Moral(p) = \frac{\prod_{X \in dom(\phi_{ij})} |\Omega_X|}{\prod_{Y \in \mathbf{W}} |\Omega_Y|} \tag{8}$$

Li and D'Ambrosio [9] also proposed an heuristic that selects a pair that minimises the score $s_1$ and maximises the score $s_2$:

$$s_1(p) = |dom(\phi_{ij})| - |\mathbf{W}| \tag{9}$$

$$s_2(p) = |dom(\phi_i)| + |dom(\phi_j)| \tag{10}$$

Any of the heuristics previously mentioned can also be used for selecting a pair of utility potentials at steps 2.*c* and 2.*b* of Definitions 3 and 5 respectively. These heuristics will be considered in the experimental analysis.

### 3.5 Example

Let us consider the ID in Fig. 1 to illustrate the behaviour of the SPI algorithm as described in this paper. In order to simplify the notation, $\phi(X_1,\ldots,X_n)$ will be denoted $\phi_{X_1,\ldots,X_n}$. First, SPI proceeds to remove variables in the chance set $\mathscr{I}_1 = \{B,C,E,F,G\}$ using the algorithm in Definition 2. The initial combination candidate set is:

$$\{\phi_C;\phi_E\},\{\phi_C;\phi_F\},\{\phi_C;\phi_G\},\{\phi_C;\phi_{BCEFG}\},\{\phi_C;\phi_{AB}\},\{\phi_E;\phi_F\},\{\phi_E;\phi_G\},\{\phi_E;\phi_{BCEFG}\},$$
$$\{\phi_E;\phi_{AB}\},\{\phi_F;\phi_G\},\{\phi_F;\phi_{BCEFG}\},\{\phi_F;\phi_{AB}\},\{\phi_G;\phi_{BCEFG}\},\{\phi_G;\phi_{AB}\},\{\phi_{BCEFG};\phi_{AB}\}$$

If the *minimum size* heuristic is used for selecting the next pair of potentials, there are 6 pairs minimizing this score. Let us suppose that the pair $\{\phi_C;\phi_E\}$ is chosen, then the resulting potential is $\phi_{CE}$. There is not any variable that can be removed, since $C$ and $E$ appear in other potentials (e.g., $\phi_{BCEFG}$). Then, the set $B$ is updated by removing pairs containing $\phi_C$ or $\phi_E$ and by adding new pairwise combinations with $\phi_{CE}$:

$$\{\phi_{CE};\phi_F\},\{\phi_{CE};\phi_G\},\{\phi_{CE};\phi_{BCEFG}\},\{\phi_{CE};\phi_{AB}\},\{\phi_F;\phi_G\},$$
$$\{\phi_F;\phi_{BCEFG}\},\{\phi_F;\phi_{AB}\},\{\phi_G;\phi_{BCEFG}\},\{\phi_G;\phi_{AB}\},\{\phi_{BCEFG};\phi_{AB}\}$$

The process will continue by choosing a pair to combine until all variables have been removed. The whole process is shown in Fig. 2 in a factor graph [7]. Nodes without any parent correspond to initial potentials while child nodes correspond to the resulting potentials of a combination. The numbers above each potentials indicate the combination ordering and arcs labels indicate the variables that are sum-marginalized.

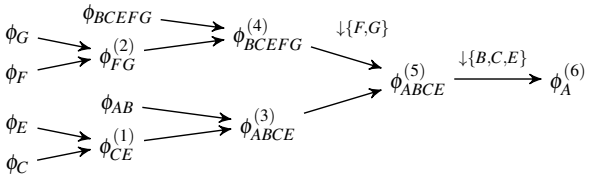



**Fig. 2.** Combination order of the probability potentials obtained using SPI for removing the chance set $\mathscr{I}_1 = \{B,C,E,F,G\}$ during the evaluation of the ID shown in Fig.1

In the $4^{th}$ iteration, after generating the potential $\phi_{BCEFG}$, variables $F$ and $G$ can be removed. Then, the algorithm in Definition 3 is executed in order to combine utility potentials and max-marginalize these variables: the combination candidate set of utility potentials is $B := \{\{\psi_{D_1FG}\}\}$ and the resulting potentials are $\phi_{BCE}$ and $\psi_{D_1BCE}$. Similarly, in the $5^{th}$ iteration, variables $B,C$ and $E$ can be removed. Now, the combination candidate set contains a pair and a singleton, that is $B := \{\{\psi_{D_1CE};\psi_{D_1BCE}\},\{\psi_{D_1BCE}\}\}$. The element selected from $B$ is the pair $\{\psi_{D_1CE};\psi_{D_1BCE}\}$. The variables $B,C$ and $E$ can be removed after adding both utility potentials in the pair, thus it is not needed to perform any additional iteration. The resulting potentials are $\phi_A$ and $\psi_{D_1A}$ which are also, in this case, the resulting potentials of algorithm in Definition 2. SPI will now proceed to remove decision $D_1$ using Definition 4 and chance variable $A$ using Definition 2.

## 4 Experimental Work

For testing the SPI algorithm, a set of 10 IDs found in the literature are used: NHL and IctNeo are two real world IDs used for medical purposes [15,16]; the oil wildcatter's problem with one and two utilities [17,18]; the Chest Clinic ID [19] obtained from the Asia BN; an ID representing the decision problem in the poker game [11]; an ID used at agriculture for treating the mildew [11]; finally, three synthetic IDs are used: the motivation example shown in Fig.1 with binary and not binary variables and the ID used by Jensen et al. in [6]. Each ID is evaluated using the SPI and the VE algorithms with the heuristics shown in Section 3.4. The *Li and D'Ambrosio* heuristic is not used with the VE algorithm because it is a specific heuristic for the SPI algorithm. An efficiency improvement used in both algorithms consists on discarding any unity probability potential generated.

Table 1 shows the total number of operations needed for each evaluation, that is the number of multiplications, divisions, additions and maximum comparisons. The ratio of the number of operations using SPI to the number of operations using VE is also shown. It can be observed that SPI requires a lower number of operations than VE in 7 out of 10 IDs when using the *minimum size* and the *Cano and Moral* heuristic. By contrast, if the *minimum weight* heuristic is used instead, SPI offers better results in 6 out of 10 IDs. Comparing *Li and D'Ambrosio* heuristics with the rest, it can be seen that this criteria only offers better results in 2 out 10 IDs.

**Table 1.** Number of operations needed for evaluating each ID using SPI and VE algorithms and different heuristics

| | *min_size* | | | *min_weight* | | | *cano_moral* | | | *li_dambrosio* |
|---|---|---|---|---|---|---|---|---|---|---|
| ID | SPI | VE | ratio | SPI | VE | ratio | SPI | VE | ratio | SPI |
| NHL | $\mathbf{2.74 \cdot 10^6}$ | $5.04 \cdot 10^6$ | 0.54 | $6.96 \cdot 10^6$ | $\mathbf{4.95 \cdot 10^6}$ | 1.41 | $\mathbf{6.96 \cdot 10^6}$ | $8.80 \cdot 10^6$ | 0.79 | $2.05 \cdot 10^7$ |
| IctNeo | $2.42 \cdot 10^6$ | $\mathbf{4.34 \cdot 10^5}$ | 5.57 | $2.36 \cdot 10^6$ | $\mathbf{3.90 \cdot 10^5}$ | 6.04 | $2.36 \cdot 10^6$ | $\mathbf{4.36 \cdot 10^5}$ | 5.4 | $1.07 \cdot 10^6$ |
| Oil Wildcatter | **125** | 150 | 0.83 | **125** | 150 | 0.83 | **125** | 157 | 0.8 | 125 |
| Oil Split Costs | **137** | 162 | 0.85 | **137** | 162 | 0.85 | **137** | 169 | 0.81 | 137 |
| Chest Clinic (Asia) | **598** | 657 | 0.91 | **598** | 625 | 0.96 | **598** | 645 | 0.93 | 682 |
| Poker | 4499 | **1775** | 2.53 | 4499 | **1831** | 2.46 | 4499 | **1582** | 2.84 | $6.12 \cdot 10^4$ |
| Mildew | $\mathbf{2.63 \cdot 10^4}$ | $3.31 \cdot 10^4$ | 0.8 | $\mathbf{2.36 \cdot 10^4}$ | $3.31 \cdot 10^4$ | 0.72 | $\mathbf{2.36 \cdot 10^4}$ | $2.60 \cdot 10^4$ | 0.91 | $5.02 \cdot 10^4$ |
| motivation binary | **324** | 511 | 0.63 | **324** | 511 | 0.63 | **324** | 513 | 0.63 | 356 |
| motivation not binary | **1753** | 6559 | 0.27 | **1753** | 2273 | 0.77 | **1753** | 1919 | 0.91 | 4597 |
| Jensen et al. | 922 | **533** | 1.73 | 922 | **533** | 1.73 | 922 | **545** | 1.69 | 746 |

## 5 Conclusions and Future Work

In this work we have described how the SPI algorithm can be used for evaluating IDs, which considers the evaluation as combinatorial factorization problem. That is, SPI tries to find an optimal order for the operations of marginalization and combination. Thus, SPI is finer grained than VE. Moreover, we also propose adapting some of the heuristics that VE uses for selecting the next pair of potentials to combine.

The experimental work shows that, in many cases, the SPI algorithm can reduce the number of operations needed to evaluate an ID compared to VE. However, SPI does not strictly dominates VE. For that reason, a line of future research could be determining which features of an ID make that SPI offers better results. The efficiency of SPI for evaluating some IDs also depends on the heuristic used, thus another line of future research could be looking for alternative heuristics. One method that improves the efficiency of the evaluation is Lazy Evaluation (LE) [6], [20], which is based on message passing in a *strong junction tree*. The SPI algorithm was already proposed as method for computing the messages in the LE of Bayesian networks [21]. Thus similar ideas could be applied for computing the messages in the LE of IDs.

**Acknowledgments.** This research was supported by the Spanish Ministry of Economy and Competitiveness under project TIN2010-20900-C04-01, the European Regional Development Fund (FEDER), the FPI scholarship programme (BES-2011-050604) and the short stay in foreign institutions scholarship EEBB-I-13-07111. The authors have also been partially supported by "Junta de Andalucía" under projects TIC-06016 and P08-TIC-03717.

## References

1. Howard, R.A., Matheson, J.E.: Influence diagram retrospective. Decision Analysis 2(3), 144–147 (2005)
2. Kjræulff, U.B., Madsen, A.L.: Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis, vol. 22. Springer (2012)
3. Shachter, R.D., Bhattacharjya, D.: Solving influence diagrams: Exact algorithms. Wiley Encyclopedia of Operations Research and Management Science (2010)
4. Shachter, R.: Evaluating influence diagrams. Operations Research, 871–882 (1986)
5. Madsen, A., Jensen, F.: Lazy evaluation of symmetric Bayesian decision problems. In: Proceedings of the 15th Conference on Uncertainty in AI, pp. 382–390. Morgan Kaufmann Publishers Inc. (1999)
6. Jensen, F., Jensen, F.V., Dittmer, S.L.: From Influence Diagrams to junction trees. In: Proceedings of the 10th International Conference on Uncertainty in AI, pp. 367–373. Morgan Kaufmann Publishers Inc. (1994)
7. Bloemeke, M., Valtorta, M.: A hybrid algorithm to compute marginal and joint beliefs in Bayesian networks and its complexity. In: Proceedings of the 14th conference on Uncertainty in AI, pp. 16–23. Morgan Kaufmann Publishers Inc. (1998)
8. Shachter, R.D., D'Ambrosio, B., Del Favero, B.: Symbolic probabilistic inference in belief networks. In: AAAI, vol. 90, pp. 126–131 (1990)
9. Li, Z., d'Ambrosio, B.: Efficient inference in Bayes networks as a combinatorial optimization problem. IJAR 11(1), 55–81 (1994)
10. D'Ambrosio, B., Burgess, S.: Some experiments with real-time decision algorithms. In: Proceedings of the 12th International Conference on Uncertainty in AI, pp. 194–202. Morgan Kaufmann Publishers Inc. (1996)
11. Jensen, F., Nielsen, T.: Bayesian networks and decision graphs. Springer (2007)
12. Kjærulff, U.: Triangulation of graphs – algorithms giving small total state space. Research Report R-90-09, Department of Mathematics and Computer Science, Aalborg University, Denmark (1990)

13. Rose, D.: A graph-theoretic study of the numerical solution of sparse positive definite systems of linear equations. Graph Theory and Computing 183, 217 (1972)
14. Cano, A., Moral, S.: Heuristic algorithms for the triangulation of graphs. In: Bouchon-Meunier, B., Yager, R.R., Zadeh, L.A. (eds.) IPMU 1994. LNCS, vol. 945, pp. 98–107. Springer, Heidelberg (1995)
15. Lucas, P., Taal, B.: Computer-based decision support in the management of primary gastric non-hodgkin lymphoma. UU-CS (1998-33) (1998)
16. Bielza, C., Gómez, M., Insua, R.S., Fernández del Pozo, J.A., Barreno García, P., Caballero, S., Sánchez Luna, M.: Ictneo system for jaundice management. Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales 92(4), 307–315 (1998)
17. Raiffa, H.: Decision analysis: Introductory lectures on choices under uncertainty (1968)
18. Dawid, P., Lauritzen, S.L., Spiegelhalter, D.J.: Probabilistic networks and expert systems: Exact computational methods for Bayesian networks. Springer (2007)
19. Goutis, C.: A graphical method for solving a decision analysis problem. IEEE Transactions on Systems, Man and Cybernetics 25(8), 1181–1193 (1995)
20. Madsen, A., Jensen, F.: Lazy propagation: a junction tree inference algorithm based on lazy evaluation. Artificial Intelligence 113(1-2), 203–245 (2004)
21. Madsen, A.: Variations over the message computation algorithm of lazy propagation. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 36(3), 636–648 (2005)

# On Causal Compositional Models: Simple Examples*

Radim Jiroušek[1,2]

[1] Faculty of management, University of Economics
Jarošovská 1117/II, 69121 Jindřichův Hradec, Czech Republic
[2] Inst. of Information Theory and Automation AV ČR
radim@utia.cas.cz

**Abstract.** The "algebraic" form of representation of probabilistic causal systems by compositional models seems to be quite useful and advantageous because of two reasons. First, decomposition of the model into its low-dimensional parts makes some of computations feasible, and, second, it appears that within these models, both conditioning and intervention can be realized as a composition of the model with a degenerated one-dimensional distribution. The syntax of these two computational processes are similar to each other; they differ just by one pair of brackets. Moreover, as it is shown in the last part of this paper on examples, it appears that these models can also cope with the problem of unobserved variables elimination.

## 1 Introduction

Each of us is interested in the relation of causation from childhood; it first enables us to answer the WHY questions, and a couple of years later, more sophisticated WHAT IF questions. Moreover, referring to any textbook we can see that practically all the knowledge is explained using causal relation. The importance of causation is visible also from the fact that from most of the articles in professional journals describing data mining applications one can see that the described research was performed with the (sometimes hidden) goal to support or to uncover some new causal relations. However, this is often misleading, because, as Pearl says in his book ([5], page 40): *The sharp distinction between statistical and causal concepts can be translated into a useful principle: behind every causal claim there must lie some causal assumption that is not discernable from the joint distribution and, hence, not testable in observational studies. Such assumptions are usually provided by humans, resting on expert judgment.* Therefore, when using causal models one should keep this fact in mind. We can construct causal

* This is an extended version of the talk presented at the 16th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty held in Mariánské Lázně, Sept. 19–22, 2013 [3]. The extension manifests itself not only in a new example and a more detailed way of presentation, but mainly in a new section on hidden variables and the respective computations of conditioning and intervention (Section 6).

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 517–526, 2014.

models only when we have a knowledge allowing us to specify causal relations, allowing us to determine what is cause and what is effect.

In this paper we present some preliminary ideas regarding application of probabilistic causal models represented in a form of compositional models. To keep the presentation as simple and informal as possible, we introduce most of the concepts just on examples. We consider only finite valued random variables that are denoted by upper case Latin characters: $X, Y, Z, W, ....$ Sets of these variables are denoted by lower case characters $(x, y, ...)$, and their probability distributions are denoted using characters of a Greek alphabet $\kappa, \lambda, \mu, \nu, \pi$. So, $\kappa(X_1, \ldots, X_n)$ denotes an $n$-dimensional probability distribution. Its $n-1$-dimensional marginal distribution is denoted by $\kappa^{-X_i}$, or, denoting $x = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$ we use also the symbol $\kappa^{\downarrow x}$. The latter symbol $\kappa^{\downarrow y}$ can be used for any $y \subseteq \{X_1, \ldots, X_n\}$.

In the above mentioned Pearl's book [5] (from which the causal model studied in this paper is taken over), one of the most important notions is the concept of an *intervention*. It means that by an external force we change the value of an intervention variable. As a simple example consider two binary variables: $A$ - alarm bell rings or not, and $S$ - smoke is or is not in a room. A smoke in the room makes the alarm bell ring. Therefore, if $\pi(A, S)$ describes the relationship between these two variables then the conditional probability $\pi(A = \textit{bell rings}|S = \textit{smoke is})$ should be close to 1, and therefore $\pi(S = \textit{smoke is}|A = \textit{bell rings}) \gg \pi(S = \textit{smoke is})$. But when considering the intervention that will be here denoted by $do(A = \textit{bell rings})$, which means that by some way or another we make the alarm bell ring (no matter whether there is a smoke in the room or not) it does not create a smoke in the room. Therefore

$$\pi(S = \textit{smoke is}|do(A = \textit{bell rings})) = \pi(S = \textit{smoke is}).$$

## 2 Causal Networks

Let us continue considering the above mentioned smoke-alarm example[1] and extend it by two other variables: $D$ - day of the week, and $B$ denoting whether Bob is at work or not. In this example, Bob is an irredeemable and undisciplined smoker whose presence at work increases probability of the appearance of the smoke in the room, and therefore also the probability of the activation of the alarm. This naturally holds only in working days when he is at work.

Such a situation can be described by a Bayesian network with a graph in Figure 1(a). This Bayesian network defines a four-dimensional probability distribution

$$\kappa(D, B, S, A) = \pi_1(D)\pi_2(B|D)\pi_3(S|B)\pi_4(A|D, S) \tag{1}$$

that represents the respective knowledge.

It is well-known from Bayesian network theory (see e.g. [1]) that without changing the resulting joint distribution $\kappa$ we can modify the graph and the system of conditional distributions so that distribution $\kappa$ is defined by another (but probabilistically equivalent) Bayesian network with a graph from Figure 1(b).

[1] For another illustrative causal model from the field of economy see [6].

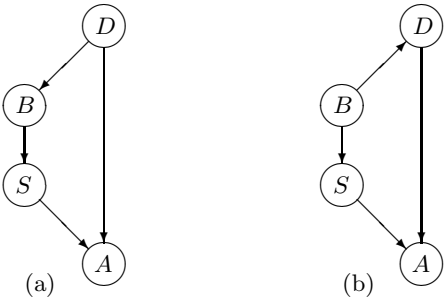



**Fig. 1.** Smoke-alarm example

However, the situation is different when we start considering the model to be causal. It means that we assume that the arrows point from the causes to effects. Now, if the graph from Figure 1(a) is causal then the graph from Figure 1(b) cannot be causal because of the different orientation of the arrows connecting nodes $D$ and $B$. Really, one can hardly imagine that asking Bob to come to work on Sunday will really change the day of week.

Generally, for causal models three types of queries are considered in Pearl's book ([5], page 29)

- **predictions** - would the alarm ring if we find the smoke in the room?
- **interventions** - would the alarm ring if we make sure that there is smoke in the room?
- **counterfactuals** - would the alarm ring had the smoke been been in the room, given that the alarm does in fact not ring and the smoke is not in the room?

In this paper we will consider the first two queries and show how they can be answered when the causal system is represented in a form of a compositional model.

## 3 Compositional Models

For a more thorough introduction to compositional model theory the reader is referred to [2]. In this paper we will introduce these models rather informally.

The basic idea is simple. Considering a three-dimensional distribution $\pi(X, Y, Z)$ and knowing that variables $X$ and $Z$ are conditionally independent given variable $Y$, one can decompose $\pi$ into its two-dimensional marginals $\pi(X, Y)$ and $\pi(Y, Z)$. It means that the original three-dimensional distribution can be unambiguously reconstructed (composed) from its two-dimensional marginals using a simple formula

$$\pi(X, Y, Z) = \frac{\pi(X, Y) \cdot \pi(Y, Z)}{\pi(Y)}.$$

This formula can be rewritten using an operator of composition $\triangleright$ that is defined as follows.

Consider two (non-empty) sets of variables $x$ and $y$. We do not impose any conditions regarding the mutual relation of these sets; they may be but need not be disjoint, one may be a subset of the other. Let $\kappa$ and $\lambda$ be distributions defined for $x$ and $y$, respectively. To avoid technical problems connected with division by zero, we assume that marginal $\lambda^{\downarrow x\cap y}$ dominates $\kappa^{\downarrow x\cap y}$, i.e.,

$$\lambda^{\downarrow x\cap y}(\cdot) = 0 \implies \kappa^{\downarrow x\cap y}(\cdot).$$

In this case we can define composition of $\kappa$ and $\lambda$ by the formula[2]

$$\kappa \triangleright \lambda = \frac{\kappa \cdot \lambda}{\lambda^{\downarrow x\cap y}}.$$

Note that for disjoint $x$ and $y$ the marginal $\kappa^{\downarrow x\cap y} = \lambda^{\downarrow x\cap y} = 1$, and $\lambda \triangleright \kappa$ simplifies to a product of (independent) distributions.

It is known that the composition of distributions $\kappa(x)$ and $\lambda(y)$ is always a distribution of variables $x \cup y$. In [2], many properties of the operator of composition are proven. In this paper we will do just with the following three that are formulated for $\kappa(x)$ and $\lambda(y)$.

$$(\kappa \triangleright \lambda)^{\downarrow x} = \kappa; \tag{2}$$
$$\kappa^{\downarrow x\cap y} = \lambda^{\downarrow x\cap y} \implies \kappa \triangleright \lambda = \lambda \triangleright \kappa; \tag{3}$$
$$z \supseteq x \cap y \implies (\kappa \triangleright \lambda)^{\downarrow z} = \kappa^{\downarrow x\cap z} \triangleright \lambda^{\downarrow y\cap z}, \tag{4}$$

from which the following two properties can easily be deduced.

$$x \supseteq y \implies \kappa \triangleright \lambda = \kappa; \tag{5}$$
$$z \subseteq x \implies \kappa^{\downarrow z} \triangleright \kappa = \kappa. \tag{6}$$

Let us start considering the iterative application of the operator of composition to a sequence of distributions. Since it is not a difficult task to show that this operator is generally neither commutative nor associative, we have to specify the ordering in which the operators are to be applied. If the opposite is not explicitly specified by brackets, we will always apply the operator of composition from left to right. Therefore, e.g.,

$$\pi \triangleright \kappa \triangleright \lambda \triangleright \mu \triangleright \nu = (((\pi \triangleright \kappa) \triangleright \lambda) \triangleright \mu) \triangleright \nu.$$

When computing the effect of intervention it will appear advantageous to compensate the lack of associativity by introducing another operator, so called

[2] To avoid technical problems, if not specified explicitly otherwise, in this paper we will consider only positive distributions. Under this assumption the dominance assumption holds for any couple of distributions and therefore their composition is always defined.

an *anticipating operator*, defining a special type of composition. For a set of variables $z$ and distributions $\kappa(x)$ and $\lambda(y)$ it is defined by the formula:

$$\kappa \circledcirc_z \lambda = (\lambda^{\downarrow(z\setminus x)\cap y} \cdot \kappa) \triangleright \lambda.$$

Its advantage is expressed by Theorem 9.4 in [2] saying that for $\pi(z)$, $\kappa(x)$ and $\lambda(y)$

$$\pi(z) \triangleright \kappa(x) \triangleright \lambda(y) = \pi(z) \triangleright (\kappa(x) \circledcirc_z \lambda(y)). \tag{7}$$

## 4 Conditioning by Composition

From now on, consider a general probability distribution $\kappa(X_1, X_2, \ldots, X_n)$ and define a *degenerated* one-dimensional probability distribution $\nu_{|i;\alpha}$ as a distribution of variable $X_i$ achieving probability 1 for value $X_i = \alpha$, i.e.,

$$\nu_{|i;\alpha}(X_i) = \begin{cases} 1 & \text{if } X_i = \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

Let us compute

$$\nu_{|i;a} \triangleright \kappa = \frac{\nu_{|i;a}(X_i) \cdot \kappa(X_1, \ldots, X_n)}{\kappa^{\downarrow\{X_i\}}}$$

for any combination of values of all variables $X_1, \ldots, X_n$. It is clear that if $X_i \neq \alpha$, $\nu_{|i;a} \triangleright \kappa = 0$. In opposite case, i.e., if $X_i = \alpha$, then

$$\begin{aligned} \nu_{|i;a} \triangleright \kappa &= \frac{\kappa(X_1, \ldots, X_{i-1}, X_i = \alpha, X_{i+1}, \ldots, X_n)}{\kappa^{\downarrow\{X_i\}}(X_i = a)} \\ &= \kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n | X_i = \alpha). \end{aligned}$$

It means that $\nu_{|i;a} \triangleright \kappa$ is an $n$-dimensional distribution that equals 0 for all combinations of values for which $X_i \neq \alpha$. In case that $X_i = \alpha$, then it equals the conditional distribution $\kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n | X_i = \alpha)$. Therefore

$$\kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n | X_i = \alpha) = (\nu_{|i;a} \triangleright \kappa)^{-X_i}.$$

Naturally, this way of expressing conditional distributions can also be used for distributions represented as compositional models. Therefore, for

$$\kappa(X_1, \ldots, X_n) = \mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m$$

we get

$$\kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n | X_i = \alpha) = \big(\nu_{|i;a} \triangleright (\mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m)\big)^{-X_i}.$$

As said above, in this paper we do not have a space to go more deeply into theory of compositional models, nevertheless let us state that the brackets in the preceding formula are important. This is because the operator of composition is not associative. Moreover, in the next section we will show an important property, namely: if $\mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m$ is a causal model then the result of intervention is computed by the formula

$$\kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n | do(X_i = \alpha)) = \big(\nu_{|i;a} \triangleright \mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m\big)^{-X_i}. \tag{8}$$

## 5 Compositional Causal Models

Consider a system $\{X_1, X_2, \ldots, X_n\}$ of finite-state variables. For each variable let $\mathcal{C}(X_i)$ denote the set of the variables that are causes of $X_i$. Naturally, some $\mathcal{C}(X_i)$ may be empty (in fact, to get a Markovian model at least one of these sets *must* be empty), and $X_i \notin \mathcal{C}(X_i)$. Using Pearl's terminology [5], we say that the causal model is *Markovian* if there exists an ordering of variables (without loss of generality we will assume that it is the ordering $X_1, X_2, \ldots, X_n$) such that $\mathcal{C}(X_1) = \emptyset$, and for all $i = 2, 3, \ldots, n$ $\mathcal{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$.

For the sake of simplicity denote $x_i = \mathcal{C}(X_i) \cup \{X_i\}$. If we have probability distributions $\mu_i(x_i)$ we can construct a compositional causal model (CCM) as

$$\kappa(X_1, \ldots, X_n) = \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n).$$

There are several theorems in [2] saying under what conditions one can change the ordering of distributions in a compositional model without influencing the resulting joint distribution. It is important to stress that for causal models, most of such transformations are forbidden. For causal models, we can consider only the orderings guaranteeing their Markovianity, i.e., for which $\mathcal{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$. And it is the result of Kratochvíl that says that all these orderings define the same joint probability distribution $\kappa(X_1, \ldots, X_n)$ (see [4]).

In what follows we will need the following three properties (for the respective proofs see [2]). For distributions $\mu_1(x_1), \mu_2(x_2), \mu_3(x_3)$

$$x_1 \supseteq x_2 \cap x_3 \Longrightarrow \mu_1 \triangleright \mu_2 \triangleright \mu_3 = \mu_1 \triangleright \mu_3 \triangleright \mu_2; \tag{9}$$

$$x_1 \supseteq x_2 \cap x_3 \Longrightarrow \mu_1 \triangleright \mu_2 \triangleright \mu_3 = \mu_1 \triangleright (\mu_2 \triangleright \mu_3); \tag{10}$$

$$x_2 \supseteq x_1 \cap x_3 \Longrightarrow \mu_1 \triangleright \mu_2 \triangleright \mu_3 = \mu_1 \triangleright (\mu_2 \triangleright \mu_3). \tag{11}$$

At the end of the previous section we promised to show how to compute the effect of an intervention in CCMs. Let us repeat the idea of Pearl [5], who computes it as a conditioning in a Bayesian network, in which all the arrows heading to the intervention node are deleted. To show that we do the same in a CCM we will need a possibility to find a causal graph corresponding to a given CCM.

Consider a CCM $\mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)$. If it is constructed in the way described at the beginning of this section then for each $i = 1, 2, \ldots, n$ the set $x_i \setminus (x_1 \cup \ldots \cup x_{i-1})$ is a singleton (i.e., $|x_i \setminus (x_1 \cup \ldots \cup x_{i-1})| = 1$). In a few lines below we will need a minor generalization of this condition, namely that

$$|x_i \setminus (x_1 \cup \ldots \cup x_{i-1})| \leq 1. \tag{12}$$

Let us construct a causal graph from $\mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)$ meeting condition (12). The elements from $x_1 \cup \ldots \cup x_n$ are nodes of the constructed causal graph, and there is an arrow $(X_i \to X_j)$ in this graph if and only if there is a distribution $\mu_k(x_k)$ in the CCM, for which

$$\begin{aligned} &\{X_i, X_j\} \subseteq x_k; \\ &X_j \notin x_1 \cup \ldots \cup x_{k-1}; \\ &X_i \in x_1 \cup \ldots \cup x_{k-1}. \end{aligned}$$

Now, consider two different CCMs:

$$\kappa(X_1,\ldots,X_n) = \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n),$$
$$\lambda(X_1,\ldots,X_n) = \kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n),$$

and construct for both of them the corresponding causal graphs. It is evident that if condition (12) holds for CCM $\kappa$, then it holds true also for CCM $\lambda$. Moreover, the reader can almost immediately see that, in the causal graph corresponding to CCM $\lambda$, there are no arrows heading to node $X_i$, and that all the other arrows from the causal graph corresponding to CCM $\kappa$ are preserved in the causal graph corresponding to CCM $\lambda$. It means that an intervention in CCM $\kappa$ can be done through conditioning in CCM $\lambda$ (see [5]):

$$\begin{aligned}\kappa(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n|do(X_i=\alpha)) &\\ = \lambda(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n|X_i=\alpha) &\\ = \left(\nu_{|i;a} \triangleright (\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n))\right)^{-X_i}&\end{aligned}$$

So, to show the validity of the expression (8) we have to show that

$$\begin{aligned}\nu_{|i;a} \triangleright \left(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)\right) &\\ = \nu_{|i;a} \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n).&\end{aligned} \tag{13}$$

When computing $\nu_{|i;a} \triangleright \left(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)\right)$, it is important to realize that both $\nu_{|i;a}$ and $\kappa^{\downarrow X_i}$ are distributions defined for the same variable $X_i$. Therefore we can apply property (11)

$$\begin{aligned}\nu_{|i;a} \triangleright \left(\left(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_{n-1}(x_{n-1})\right) \triangleright \mu_n(x_n)\right) &\\ = \nu_{|i;a} \triangleright \left(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_{n-1}(x_{n-1})\right) \triangleright \mu_n(x_n).&\end{aligned}$$

The same idea can also be applied to a shorter sequence, which yields

$$\begin{aligned}\nu_{|i;a} \triangleright \left(\left(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_{n-2}(x_{n-2})\right) \triangleright \mu_{n-1}(x_{n-1})\right) &\\ = \nu_{|i;a} \triangleright \left(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_{n-2}(x_{n-2})\right) \triangleright \mu_{n-1}(x_{n-1}).&\end{aligned}$$

Thus, applying property (11) $n$ times we get that

$$\begin{aligned}\nu_{|i;a} \triangleright \left(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)\right) &\\ = \nu_{|i;a} \triangleright \kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n),&\end{aligned}$$

and to show the validity of the required equation (13) it is enough to apply property (5) to the first operator of composition.

## 6 Causal Models with Unobserved Variables

Consider a simple Markovian causal model that is treated in Section 3.4.3 of [5]. It has four variables $U, X, Y$ and $Z$, where the first variable is unobserved (latent, unmeasured) whilst all the others are measurable. The causal structure of the considered model is defined by the sets $\mathcal{C}(U) = \emptyset$, $\mathcal{C}(X) = \{U\}$, $\mathcal{C}(Y) = \{X\}$, and $\mathcal{C}(Z) = \{U, Y\}$. Thus, this causal model is represented by only one CCM, namely

$$\kappa(U) \triangleright \kappa(U, X) \triangleright \kappa(X, Y) \triangleright \kappa(U, Y, Z). \tag{14}$$

However, since variable $U$ is unobservable, there is no way how to estimate the necessary marginal distributions $\kappa(U), \kappa(U, X)$ and $\kappa(U, Y)$, we can get only estimates of $\kappa(X, Y, Z)$ (and its maginals). This, as a matter of course, enables us to compute "predictive" conditional distributions like $\kappa(Y|X = \alpha)$ and $\kappa(Z|X = \alpha)$. Nevertheless, there is a natural question whether there is a chance to compute also the effects of interventions $\kappa(Y|do(X = \alpha))$, $\kappa(Z|do(Y = \beta))$, $\kappa(Z|do(X = \alpha))$, $\kappa(Y, Z|do(X = \alpha))$ and $\kappa(X, Z|do(Y = \beta))$. In [5] Pearl computes all the above mentioned effects and says that *in all the derivations, the graph* (of the respective causal network) *provides both the license for applying the inference rules and the guidance for choosing the right rule to apply*. Surprisingly enough, thanks to the compositional form of the model we can do the necessary computations just with the help of the above introduced properties.

A trivial solution is for $\kappa(Y|do(X = \alpha))$[3]:

$$\begin{aligned}
\kappa(Y|do(X = \alpha)) &= \big(\nu_{|X;\alpha} \triangleright \kappa(U) \triangleright \kappa(U, X) \triangleright \kappa(X, Y) \triangleright \kappa(U, Y, Z)\big)^{\downarrow\{Y\}} \\
&\overset{(5)}{=} \Big(\big(\nu_{|X;\alpha} \triangleright \kappa(U) \triangleright \kappa(X, Y) \triangleright \kappa(U, Y, Z)\big)^{-\{Z\}}\Big)^{\downarrow\{Y\}} \\
&\overset{(4)}{=} \Big(\nu_{|X;\alpha} \triangleright \kappa(U) \triangleright \kappa(X, Y) \triangleright \kappa(U, Y)\Big)^{\downarrow\{Y\}} \\
&\overset{(5)}{=} \Big(\big(\nu_{|X;\alpha} \triangleright \kappa(U) \triangleright \kappa(X, Y)\big)^{-\{U\}}\Big)^{\downarrow\{Y\}} \\
&\overset{(4)}{=} \big(\nu_{|X;\alpha} \triangleright \kappa(X, Y)\big)^{\downarrow\{Y\}} = \kappa(Y|X = \alpha),
\end{aligned}$$

which corresponds to our intuition.

A bit more complicated is to compute

$$\begin{aligned}
\kappa(Z|do(Y = \beta)) &= \big(\nu_{|Y;\beta} \triangleright \kappa(U) \triangleright \kappa(U, X) \triangleright \kappa(X, Y) \triangleright \kappa(U, Y, Z)\big)^{\downarrow\{Z\}} \\
&\overset{(11)}{=} \big(\nu_{|Y;\beta} \triangleright (\kappa(U) \triangleright \kappa(U, X)) \triangleright \kappa(X, Y) \triangleright \kappa(U, Y, Z)\big)^{\downarrow\{Z\}} \\
&\overset{(6)}{=} \big(\nu_{|Y;\beta} \triangleright \kappa(U, X) \triangleright \kappa(X, Y) \triangleright \kappa(U, Y, Z)\big)^{\downarrow\{Z\}} \\
&\overset{(6)}{=} \big(\nu_{|Y;\beta} \triangleright (\kappa(X) \triangleright \kappa(U, X)) \triangleright \kappa(X, Y) \triangleright \kappa(U, Y, Z)\big)^{\downarrow\{Z\}} \\
&\overset{(11)}{=} \big(\nu_{|Y;\beta} \triangleright \kappa(X) \triangleright \kappa(U, X) \triangleright \kappa(X, Y) \triangleright \kappa(U, Y, Z)\big)^{\downarrow\{Z\}}
\end{aligned}$$

[3] The numbers above the signs of equation refer to the properties used.

$$
\begin{aligned}
&\overset{(10)}{=} \big(\nu_{|Y;\beta} \triangleright \kappa(X) \triangleright (\kappa(U,X) \triangleright \kappa(X,Y)) \triangleright \kappa(U,Y,Z)\big)^{\downarrow\{Z\}}\\
&\overset{(11)}{=} \big(\nu_{|Y;\beta} \triangleright \kappa(X) \triangleright (\kappa(U,X) \triangleright \kappa(X,Y) \triangleright \kappa(U,Y,Z))\big)^{\downarrow\{Z\}}\\
&= \Big(\big(\nu_{|Y;\beta} \triangleright \kappa(X) \triangleright \kappa(U,X,Y,Z)\big)^{-U}\Big)^{\downarrow\{Z\}}\\
&\overset{(4)}{=} \big(\nu_{|Y;\beta} \triangleright \kappa(X) \triangleright \kappa(X,Y,Z)\big)^{\downarrow\{Z\}}\\
&= \big(\kappa(X) \triangleright \nu_{|Y;\beta} \triangleright \kappa(X,Y,Z)\big)^{\downarrow\{Z\}}.
\end{aligned}
$$

It is more sophisticated to compute $\kappa(Z|do(X=\alpha))$. For this we will use, among others, the anticipating operator (introduced in Section 3).

$$
\begin{aligned}
\kappa(Z|do(X=\alpha)) &= \big(\nu_{|X;a} \triangleright \kappa(U) \triangleright \kappa(U,X) \triangleright \kappa(X,Y) \triangleright \kappa(U,Y,Z)\big)^{\downarrow\{Z\}}\\
&\overset{(5)}{=} \big(\nu_{|X;a} \triangleright \kappa(U) \triangleright \kappa(X,Y) \triangleright \kappa(U,Y,Z)\big)^{\downarrow\{Z\}}\\
&\overset{(9)}{=} \big(\nu_{|X;a} \triangleright \kappa(X,Y) \triangleright \kappa(U) \triangleright \kappa(U,Y,Z)\big)^{\downarrow\{Z\}}\\
&\overset{(7)}{=} \Big(\nu_{|X;a} \triangleright \kappa(X,Y) \triangleright \Big(\kappa(U) \circledcirc_{\{X,Y\}} \kappa(U,Y,Z)\Big)\Big)^{\downarrow\{Z\}}\\
&\overset{(4)}{=} \Big(\nu_{|X;a} \triangleright \kappa(X,Y) \triangleright \Big(\kappa(U) \circledcirc_{\{X,Y\}} \kappa(U,Y,Z)\Big)^{-U}\Big)^{\downarrow\{Z\}}
\end{aligned}
$$

To express $\Big(\kappa(U) \circledcirc_{\{X,Y\}} \kappa(U,Y,Z)\Big)^{-U}$ we will use the idea of extension used for this purpose by Pearl in [5].

$$
\begin{aligned}
&\Big(\kappa(U) \circledcirc_{\{X,Y\}} \kappa(U,Y,Z)\Big)^{-U} = \Big(\kappa(U) \circledcirc_{\{Y\}} \kappa(U,Y,Z)\Big)^{-U}\\
&= \Big(\Big(\kappa(U,X) \circledcirc_{\{Y\}} \kappa(U,Y,Z)\Big)^{-X}\Big)^{-U} = \big(\lambda(Y) \cdot \kappa(U,X) \triangleright \kappa(U,Y,Z)\big)^{\downarrow\{Y,Z\}}\\
&= \big(\lambda(Y) \cdot \kappa(X) \triangleright \kappa(U,X) \triangleright \kappa(U,Y,Z)\big)^{\downarrow\{Y,Z\}}\\
&\overset{(5)}{=} \big(\lambda(Y) \cdot \kappa(X) \triangleright \kappa(U,X) \triangleright \kappa(X,Y) \triangleright \kappa(U,Y,Z)\big)^{\downarrow\{Y,Z\}}\\
&\overset{(10)}{=} \big(\lambda(Y) \cdot \kappa(X) \triangleright (\kappa(U,X) \triangleright \kappa(X,Y)) \triangleright \kappa(U,Y,Z)\big)^{\downarrow\{Y,Z\}}\\
&= \big(\lambda(Y) \cdot \kappa(X) \triangleright \kappa(U,X,Y) \triangleright \kappa(U,Y,Z)\big)^{\downarrow\{Y,Z\}}
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(11)}{=} \big(\lambda(Y) \cdot \kappa(X) \triangleright (\kappa(U,X,Y) \triangleright \kappa(U,Y,Z))\big)^{\downarrow\{Y,Z\}}\\
&= \big(\lambda(Y) \cdot \kappa(X) \triangleright \kappa(U,X,Y,Z)\big)^{\downarrow\{Y,Z\}} \overset{(4)}{=} \big(\lambda(Y) \cdot \kappa(X) \triangleright \kappa(X,Y,Z)\big)^{\downarrow\{Y,Z\}}\\
&= \Big(\kappa(X) \circledcirc_{\{Y\}} \kappa(X,Y,Z)\Big)^{-X},
\end{aligned}
$$

which eventually leads to

$$\kappa(Z|do(X=\alpha)) = \left(\nu_{|X;a} \triangleright \kappa(X,Y) \triangleright \left(\kappa(X) \circledcirc_{\{Y\}} \kappa(X,Y,Z)\right)^{-X}\right)^{\downarrow\{Z\}}.$$

## 7 Conclusions

As promised in Introduction, we have presented some preliminary ideas regarding causal compositional models. Namely, we have shown how to compute intervention in these models. For the sake of simplicity we have assumed that the considered probability distributions are positive. Let us stress that this assumption can easily be replaced by the assumption of dominance of the respective distributions. So, the approach can be used also in situations when some of the considered dependencies are deterministic.

The computations presented in the last section may seem rather tedious and laborious. This is because we proceeded very slowly making one elementary modification at each step. In fact, it is rather surprising that as for tools from compositional model theory, we could do just with a small battery of elementary rules, most of which can easily be deduced just from the definition of the operator of composition. The only step that can be considered as advanced was connected with the application of the anticipating operator.

**Acknowledgement.** The research was partially supported by GA ČR under the project number GA 403/12/2175.

## References

1. Jensen, F.V., Nielsen, T.D.: Bayesian networks and decision graphs. Springer, Heidelberg (2007)
2. Jiroušek, R.: Foundations of compositional model theory. Int. J. General Systems 40(6), 623–678 (2011)
3. Jiroušek, R.: On Causal Compositional Models: preliminaries. In: Proceedings of the 16th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty. FM VŠE, Jindřův Hradec, pp. 73–81 (2013)
4. Kratochvíl, V.: Probabilistic Compositional Models: solution of an equivalence problem. Int. J. Approximate Reasoning 54(5), 590–601 (2013)
5. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press, New York (2009)
6. Váchová, L., Bína, V.: A Note on Computation of Causal Effect in Economic Models. In: Proceedings of the 31st Mathematical models in Economy, pp. 974–979 (2013)

# How to Create Better Performing Bayesian Networks: A Heuristic Approach for Variable Selection

Esma Nur Cinicioglu* and Gülseren Büyükuğur

Istanbul University, School of Business, Quantitative Methods Division,
Avcilar, 34322, Istanbul, Turkey
esmanurc@istanbul.edu.tr, gulsayici@gmail.com

**Abstract.** Variable selection in Bayesian networks is necessary to assure the quality of the learned network structure. Cinicioglu & Shenoy (2012) suggested an approach for variable selection in Bayesian networks where a score, $S_j$, is developed to assess each variable whether it should be included in the final Bayesian network. However, with this method the variables without parents or children are punished which affects the performance of the learned network. To eliminate that drawback, in this paper we develop a new score, $NS_j$. We measure the performance of this new heuristic in terms of the prediction capacity of the learned network, its lift over marginal and evaluate its success by comparing it with the results obtained by the previously developed $S_j$ score. For the illustration of the developed heuristic and comparison of the results credit score data is used.

**Keywords:** Bayesian networks, Variable selection, Heuristic.

## 1 Introduction

The upsurge of popularity of Bayesian networks brings a parallel increase in research for structure learning algorithms of Bayesian networks from data sets. The ability of Bayesian networks to represent the probabilistic relationships between the variables is one of the main reasons of the rise in reputation of Bayesian networks as an inference tool. This also generates the major appeal of Bayesian networks for data mining. With the advancement and diversification of the structure learning algorithms, more variables may be incorporated to the learning process, bigger data sets may be used for learning, and inferences become faster even in the presence of continuous variables. The progress achieved on structure learning algorithms for Bayesian networks is encouraging for the increasing use of Bayesian networks as a general decision support system, a data mining tool and for probabilistic inference. On the other hand, though the quality of a learned network may be evaluated by many different aspects, the performance of the learned network very much depends on the selection of the variables to be included to the network. Depending on the purpose of the application, the characteristics of an application may differ and hence the expectations from a Bayesian network performance may vary. Therefore to assure to end up with a Bayesian

* Corresponding author.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 527–535, 2014.

network of high quality, variable selection in Bayesian networks should constitute an important dimension of the learning process. There is a considerable literature in statistics on measures like AIC, BIC, Caplow's C-p statistic etc. that are used for variable selection in statistical models. These measures have been adopted by the machine learning community for evaluating the score based methods for learning Bayesian network models (Scutari, 2010). However, these scores are used as a measure of the relative quality of the learned network and do not assist in the variable selection process. Additionally, as discussed in Cui *et al.* (2010), traditional methods of stepwise variable selection do not consider the interrelations among variables and may not identify the best subset for model building. Despite the interest for structure learning algorithms and adaptation of different measures for the evaluation of the resulting Bayesian networks, variable selection in Bayesian networks is a topic which needs further attention of the researchers. Previously Koller and Sahmi (1996) elaborate the importance of feature selection and state that the goal should be to eliminate a feature if it gives us little or no additional information. Hruschka *et al.* (2004) described Bayesian feature selection approach for classification problems. In their work, first a BN is created from a dataset and then the Markov blanket of the class variable is used to the feature subset selection task. Sun & Shenoy (2007) provided a heuristic method to guide the selection of variables in naïve Bayes models. To achieve the goal, the proposed heuristic relies on correlations and partial correlations among variables. Another heuristic developed for variable selection in Bayesian networks was proposed by Cinicioglu & Shenoy (2012). With this heuristic a score called $S_j$ was developed which helps to determine the variables to be used in the final Bayesian network. By this heuristic first an initial Bayesian network is developed with the purpose of learning the conditional probability tables (cpts) of all the variables in the network. The cpts indicate the association of a variable with the other variables in the network. Using the cpt of each variable, its corresponding $S_j$ score is calculated. In their paper Cinicioglu & Shenoy (2012) illustrate that by applying proposed heuristic the performance of the learned network in terms of the prediction capacity may be improved substantially. In this paper first we discuss the $S_j$ score, and then identify the problem that though the $S_j$ score demonstrates a sound performance on prediction capacity, its formula leads to the problem that the variables without parents or children in the network are punished and that in turn affects the overall performance of the heuristic. Trying to eliminate that drawback, in this paper we suggest a modified version of the $S_j$ score, which is called as $NS_j$. We measure the performance of this new score in terms of the prediction capacity of the learned network, its lift compared to the marginal model and evaluate its success by comparing it with the results obtained by the previously developed $S_j$ score. For the illustration of the developed heuristic and comparison of the results credit score data is used. The outline of the remainder of the paper is as follow: The next section gives details about the credit data set used for the application of the proposed heuristic. In section 3 the development of the new heuristic is explained, where both $S_j$ and $NS_j$ scores are discussed in detail in subsections 3.1 and 3.2 respectively. In section 4, using both of the variable selection scores $S_j$ and $NS_j$, different Bayesian networks are created. The performance results of these two heuristics are compared in terms of the prediction capacity and improvement rates obtained compared to the marginal model.

## 2 Data Set

The data set used in this study is a free data set, called the German credit data, provided by the UCI Center for Machine Learning and Repository Systems. The original form of the data set contains the information of 1000 customers about 20 different attributes, 13 categorical and 7 numerical, giving the information necessary to evaluate a customer's eligibility to get a credit. Before the use of the data set for the application of the proposed heuristics several changes are made in the original data set. In this research, the German credit data set is transformed into a form where the numerical attributes "Duration in month", "Credit amount", "Installment rate in percentage of disposable income", "Present residence since", "Age in years", "Number of existing credits at this bank" and "Number of people being liable to provide maintenance for" are discretized. The variable "Personal status and sex" is divided into two categorical variables as "Personal status" and "Sex". In the original data set the categorical variable "Purpose" contains eleven different states. In this paper some of these states are joined together, like "car" and "used car" as "car", "furniture", "radio" and "domestic appliances" as "appliances" and "retraining" and "business" as "business", resulting in seven different states at the end. The final data set used in this study constitutes of 21 columns and 1000 lines, referring the number of variables and cases consequently.

## 3 Development of the Proposed Heuristic

### 3.1 $S_j$ Score

The heuristic developed by Cinicioglu & Shenoy (2012) is based on the principle that a good prediction capacity of a Bayesian network depends on the choice of the variables that have high associations with each other. A marginal variable present in a network will not have any dependencies with the remaining variables in the network and thus won't have any impact for the overall performance of the network. In that instance, the arcs learned using an existing structure learning algorithm shows the dependency of a child node with its parent node, hence a proof of association. However, not all variables which do not place themselves as marginals, can be incorporated to the final Bayesian network. The idea is to develop an efficient heuristic for variable selection where the Bayesian network created using the selected variables will show a superior prediction performance compared to the random inclusion of variables to the network. Besides, though the presence of an arc shows the dependency relationship between two variables in the network, the degree of association is not measured there and may vary quite differently among variables. A natural way to examine the association of a variable with other variables considered for inclusion in the final Bayesian network is to learn an initial Bayesian network structure at first and then use the conditional probability tables of each variable as a source of measurement for the degree of association.

Applying the distance measure to the conditional probability table of a variable, the degree of change on the conditional probabilities of a child node depending on the states of its parents may be measured. In that instance a high average distance obtained indicate that the conditional probability of the variable considered changes a great deal depending on the states of its parents. Thus, a high average distance is an indication of the high association of a child node with its parents.

The average distance of each variable may be calculated using the formula given below. Here $d$ represents the average distance of the variable of interest with its parent variables. $p$ and $q$ stand for the conditional probabilities of this variable for the different states of its parents, $i$ stands for the different states of the child node and $n$ stands for the number of states of the set of parent nodes.

$$d = \sum(p_i - q_i)^2 / \binom{n}{2}, \quad \forall pq \tag{1}$$

However, there may be variables in the network which do not have a high level of association with its parent node but do possess a high association with its children. Basing the selection process on the average distance of each variable solely will deteriorate the performance of the network created. Besides, while the average distance obtained from the cpt of a variable shows the degree of association of a child node with its parents, the same average distance also shows the degree of association of a parent node with its child, jointly with the child's other parents. Following this logic Cinicioglu & Shenoy (2012) developed the $S_j$ score given in Equation (2) below. In this formula the $S_j$ score of a variable $j$ is the sum of the average distance of this variable $d_j$ and the average of the average distances of its children. Here $ij$ denotes the child variable $i$ of the variable $j$ and $c_j$ denotes the number of $j$'s children.

$$S_j = d_j + (\sum_i \frac{d_{ij}}{c_j}) \tag{2}$$

Consider Table 1 given below. This table is the cpt of the variable "Credit amount". Using the formula given in Equation (1) the average distance of this variable is calculated as 0.0107. Considering Figure 1 given below, we see that "Credit Amount" possesses three children. Hence in order to calculate the $S_j$ score of "Credit Amount" we need to find the average distances of the child variables, average them and then add it to the average distance of the "Credit Amount".

A high $S_j$ score is desired as an indication of the high association with other variables. Ideally, according to the heuristic, the variable with the lowest $S_j$ score will be excluded from the analysis and a new BN will be created with the remaining variables. This network will include the new cpts which will be the basis for the selection of the variable to be excluded from the network. This process is repeated until the desired number of variables is obtained. This repeated process is the ideal way of applying the heuristic, however if not automated will require a great deal of time. In the following, subsection 3.2, the shortcomings of the $S_j$ score are discussed. As a modification of the $S_j$ score to handle the problems involved with the old variable selection method, a new score called $NS_j$ is suggested.

**Table 1.** Cpt of the variable "Credit Amount"

| | **Credit Amount** | | | | |
|---|---|---|---|---|---|
| **Telephone** | 0 - 4000 | 4000 -8000 | 8000 -12000 | 12000-16000 | 16000-20000 |
| None | 0.8286 | 0.1364 | 0.0300 | 0.0033 | 0.0017 |
| Yes | 0.6308 | 0.2347 | 0.0807 | 0.0489 | 0.0049 |

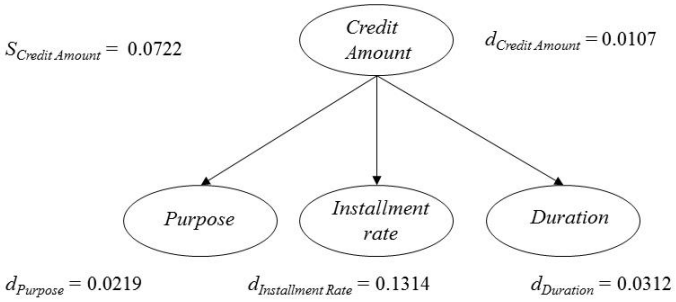



**Fig. 1.** Variable "Credit Amount" with its three children and calculation of $S_{Credit\ Amount}$

### 3.2 A New Variable Selection Score: $NS_j$

The heuristic developed by Cinicioglu & Shenoy (2012) tries to identify the variables which possess a high level of association with their parent and child variables. With that purpose the variable selection score developed, $S_j$, is comprised of two parts: $S_j$ is the sum of the average distance of the variable of interest and the average of the average distances of its children. This way, with the $S_j$ score the variable is evaluated by considering both the association with its parents and also with its children. However, this approach also has the drawback that the variables without parents or children are penalized for inclusion to the final Bayesian network.

Consider the formula of the $S_j$ score given in Equation (2). A variable without parents will only have a marginal probability distribution, not a cpt, and thus its average distance will be considered as zero. Similarly, for a variable which does not have any children the $S_j$ score will be equal to its average distance. The resulting $S_j$ scores for a variable without parents and for a variable without children are given in Equations (3) and (4) respectively.

For a variable $j$ without parents → $S_j = \sum_i \frac{d_{ij}}{c_j}$ (3)

For a variable $j$ without children → $S_j = d_j$ (4)

As illustrated above because of the formulation of the $S_j$ score, variables which do not possess parents or children will be punished in the variable selection process. If such a variable which lacks parents or children has a strong association with the present part

(its parents or children depending on the case) though, then this selection process may cause to create networks with lower performance. To overcome this problem, in this research, a modified version of the $S_j$ score, $NS_j$, is presented. For variables which lack either parents or children the score will remain to be the same as the old one. For variables which possess both parents and children on the other hand, $NS_j$ will be equal to the half of the old $S_j$ score. These two cases are formulated in Equation (5) and (6) given below.

For a variable *j* without parents or children

$$NS_j = d_j + (\sum_i \frac{d_{ij}}{c_j}) \tag{5}$$

For a variable *j* both with parents and children

$$NS_j = \frac{d_j + (\sum_i \frac{d_{ij}}{c_j})}{2} \tag{6}$$

The variables which don't have any parents or children will be eliminated from the network. In the following section both of these heuristics will be used to learn BNs from the credit data set introduced in Section 2, their performance will be evaluated in terms of the prediction capacity and improvement obtained compared to the marginal model.

## 4 Evaluation of the Proposed Heuristic

In this section the performance of the variable selection scores $S_j$ and $NS_j$ are compared. The evaluation is made in terms of the prediction capacity and improvement of the BNs created using the suggested scores. For the application of the heuristic, first, it is necessary to learn an initial BN from the data set. For illustration and evaluation of the suggested scores the credit data set given in Section 2 will be used. For learning BNs from the data set WinMine, software (Heckerman et al., 2000) developed by Microsoft Research, is used. The main advantage of WinMine is its ability to automatically calculate log-scores and lift over marginals of the learned BNs. Log-score is a quantitative criterion to compare the quality and performance of the learned BNs. The formula for the log score is given below.

$$LogScore(x_1, \ldots, x_N) = \sum_{i=1}^{N} log_2 p(x_i|model)/nN \tag{7}$$

where *n* is the number of variables, and *N* is the number of cases in the test set.

For the calculation of the log-score, the dataset is divided into a 70/30 train and test split[1] and the accuracy of the learned model on the test set is then evaluated using the log score. Using WinMine the difference between the log scores of the provided

[1] In WinMine only the percentage of the test/training test data may be determined. Using a different software in further research 10-fold cross validation will increase the validity of the results.

model and the marginal model can also be compared which is called as the "lift over marginal". A positive difference signifies that the model out-performs the marginal model on the test set. The initial BN learned from the credit data set is given in Figure 2 below.

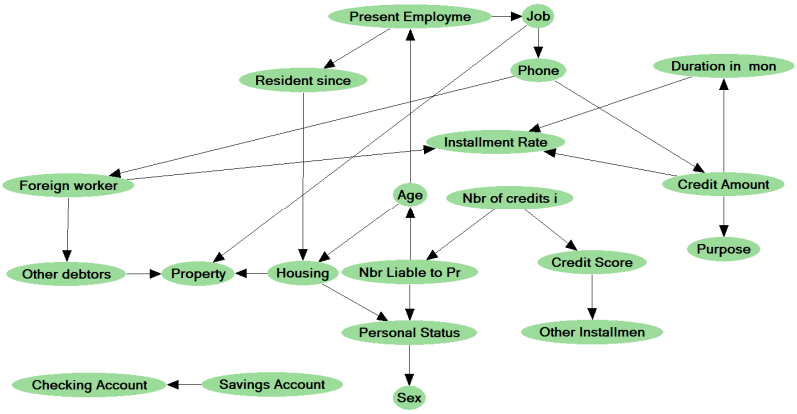



**Fig. 2.** The initial BN learned from the credit data set containing all of the variables

Using the cpts obtained through the initial BN we can calculate both the $S_j$ and $NS_j$ scores. Figure 3 given below depicts the graph of both $S_j$ and $NS_j$ scores for the 21 variables used in the initial BN. The observations made are as follows: For seven variables in the network the corresponding $S_j$ and $NS_j$ scores agree. These seven variables are the ones which either lack parents or children.

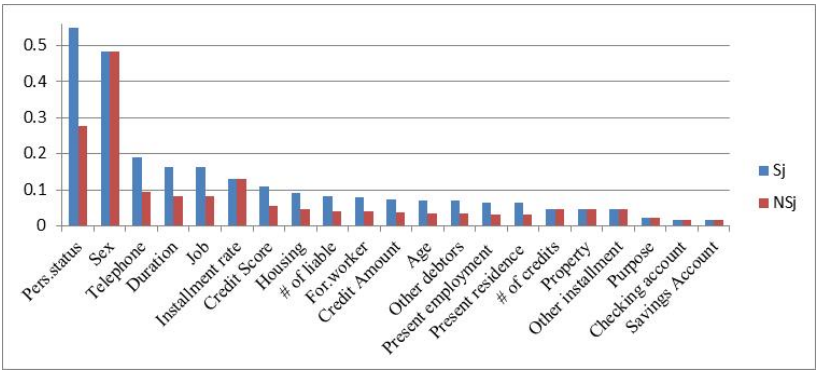



**Fig. 3.** Graph of the $S_j$ and $NS_j$ scores calculated using the cpts obtained from the initial BN

In our analysis we want to compare the performance of these two variable selection scores. With that purpose two sets of variables are created, one by selecting the variables with the highest $S_j$ scores and the second with the highest $NS_j$ scores. Using the variables selected the corresponding BNs are learned. The performance of the BNs

are compared in terms of prediction capacity of the provided model and in terms of the improvement obtained. As the next step the same process is repeated by using the cpts of the new BNs to calculate the new $S_j$ and $NS_j$ scores. Accordingly, the variables to be excluded from the network is decided according their ranking on the variable selection score considered, $S_j$ or $NS_j$. In our analysis, we repeated the steps five times and created BNs using 17, 15, 13, 11 and 8 variables, all selected according their ranking in the corresponding variable selection scores. The results of their performance are listed in Table 2 given below. Both of the variable selection scores obtain better results compared to the marginal model and also the average distance measure. Notice that also the results of the BNs created using the average distance $d_j$ are listed in the same table. This is done for comparison purposes to illustrate that both of the variable selection scores do result in superior performance compared to the average distance measure. Additionally, in almost all the networks considered except the BN with 17 variables, we obtained better performing networks using the $NS_j$ score both in terms of the prediction capacity and improvement obtained.

**Table 2.** Performance results of the variable selection scores $S_j$ and $NS_j$[2]

| | | LogScore | Prediction rate | Lift Over Marginal | Improvement obtained |
|---|---|---|---|---|---|
| **initial BN** | | – 0.76 | 59.13% | 0.19 | 7.13% |
| **Top 17 variables** | $d_j$ | – 0.83 | 56.38% | 0.17 | 6.30% |
| | $S_j$ | – 0.73 | 60.30% | 0.21 | 8.20% |
| | $NS_j$ | – 0.77 | 58.58% | 0.19 | 7.36% |
| **Top 15 variables** | $d_j$ | – 0.77 | 58.46% | 0.19 | 7.18% |
| | $S_j$ | – 0.72 | 60.87% | 0.22 | 8.48% |
| | $NS_j$ | – 0.72 | 60.87% | 0.22 | 8.48% |
| **Top 13 variables** | $d_j$ | – 0.78 | 58.29% | 0.21 | 7.84% |
| | $S_j$ | – 0.73 | 60.23% | 0.20 | 7.65% |
| | $NS_j$ | – 0.66 | 63.27% | 0.22 | 8.95% |
| **Top 11 variables** | $d_j$ | – 0.73 | 60.39% | 0.19 | 7.35% |
| | $S_j$ | – 0.72 | 60.66% | 0.19 | 7.41% |
| | $NS_j$ | – 0.65 | 63.74% | 0.22 | 9.06% |
| **Top 8 variables** | $d_j$ | – 0.76 | 58.87% | 0.18 | 6.97% |
| | $S_j$ | – 0.67 | 62.65% | 0.18 | 7.44% |
| | $NS_j$ | – 0.66 | 63.14% | 0.22 | 9.08% |

## 5 Results, Conclusions and Further Research

In order to ensure the prediction capacity of a BN learned from the data set and to be able to discover hidden information inside a big data set it is necessary to select the right set of variables to be used in the BN to be learned. This problem is especially

[2] The results are rounded to two decimal places.

apparent when there is a huge set of variables and the provided data is limited. In the last decade the research on structure learning algorithms for BNs have grown substantially. Though, there exists a wide research for variable selection in statistical models, the research conducted for variable selection in BNs remains to be limited. The variable selection measures developed for statistical models have been adapted by the machine learning community for evaluating the overall performance of the BN and do not provide guidance in variable selection for creating a good performing BN. The variable selection score $S_j$ (Cinicioglu &Shenoy, 2012), provides a sound performance for prediction capacity of the resulting network, however has the drawback that the variables without parents or children punished for inclusion to the network. Motivated by that problem in this research we suggest a modification to the $S_j$ score, called as $NS_j$ which fixes the problems inherent in its predecessor $S_j$. A credit score data set is used for applying the proposed heuristic. The performance of the resulting BNs using the proposed heuristic is evaluated using logscore and lift over marginal which provides the prediction capacity of the network and the improvement obtained using the provided model compared to the marginal model. These results are compared with the results obtained using the distance measure and the $S_j$ score. Accordingly, the new developed $NS_j$ score show better performance both in terms of prediction capacity and the improvement obtained. For further research, different variable selection scores from statistical models and different data sets may be used to evaluate the results of the proposed heuristic.

**Acknowledgements.** We are grateful for two anonymous reviewers of IPMU-2014 for comments and suggestions for improvements. This research was funded by Istanbul University Research fund, project number 27540.

## References

1. Cinicioglu, E.N., Shenoy, P.P.: A new heuristic for learning Bayesian networks from limited datasets: a real-time recommendation system application with RFID systems in grocery stores. Annals of Operations Research, 1–21 (2012)
2. Cui, G., Wong, M.L., Zhang, G.: In Bayesian variable selection for binary response models and direct marketing forecasting. Expert Systems with Applications 37, 7656–7662 (2010)
3. Heckerman, D., Chickering, D.M., Meek, C., Rounthwaite, R., Kadie, C.: Dependency Networks for Inference, Collaborative Filtering, and Data Visualization. Journal of Machine Learning Research 1, 49–75 (2000)
4. Hruschka Jr., E.R., Hruschka, E.R., Ebecken, N.F.F.: Feature selection by bayesian networks. In: Tawfik, A.Y., Goodwin, S.D. (eds.) Canadian AI 2004. LNCS (LNAI), vol. 3060, pp. 370–379. Springer, Heidelberg (2004)
5. Koller, D., Sahami, M.: Toward optimal feature selection (1996)
6. Murphy, P.M., Aha, D.W.: UCI Repository of Machine Learning Databases. Department of Information and Computer Science, University of California, Irvine, CA (1994)
7. Sun, L., Shenoy, P.P.: Using Bayesian networks for bankruptcy prediction: some methodological issues. European Journal of Operational Research 180(2), 738–753 (2007)

# A Highly Automated Recommender System Based on a Possibilistic Interpretation of a Sentiment Analysis

Abdelhak Imoussaten[1,⋆], Benjamin Duthil[2], François Trousset[1], and Jacky Montmain[1]

[1] LGI2P, Ecole des Mines d'Alès,
Site EERIE Parc Scientifique G. Besse, 30035 Nîmes Cedex, France
[2] L3i, Université de La Rochelle,
Avenue M. Cré peau, 17042 La Rochelle cedex 01, France
{abdelhak.imoussaten,francois.trousset,jacky.montmain}@mines-ales.f,
benjamin.duthil@univ-lr.fr

**Abstract.** This paper proposes an original recommender system (RS)[1] based upon an automatic extraction of trends from opinions and a multicriteria multi actors assessment model. Our RS tries to optimize the use of the available information on the web to reduce as much as possible the complex and tedious steps for multicriteria assessing and for identifying users' preference models. It may be applied as soon as i) overall assessments of competing entities are provided by trade magazines and ii) web users' critics in natural languages and related to some characteristics of the assessed entities are available. Recommendation is then based on the capacity of the RS to associate a web user with a trade magazine that conveys the same values as the user and thus represents a reliable personalized source of information. Possibility theory is used to take account subjectivity of critics. Finally a case study concerning movie recommendations is presented.

**Keywords:** Possibility theory, Intervals merging, Multicriteria aggregation, Recommender system, Opinion-mining.

## 1 Introduction

In recent years, many companies and web sites have set up systems to analyze the preferences of their users in order to better meet their expectations. To date, recommendation systems are present in many areas such as tourism /leisure, advertising, e-commerce, movies, etc. Due to the exponential growth of the quantity of data available on the Internet in recent years, searching and finding products, services and relevant contents become a difficult task for the user often drowned out by the mass of information. This explains the growing interest in recommendation systems (RS) both by users as by commercial sites.

⋆ Corresponding author.

[1] This work is an extension of [1].

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 536–545, 2014.

The task of recommendation has been identified as a way to help users to find information, or elements that are likely of interest. Roughly speaking, we consider a set of users and a set of items (products or services) that may be recommended to each user. In addition, a multicriteria recommendation improves the quality of a RS because it makes explicit the characteristics for which an item was proposed to the user [2] [3]. A RS that takes advantage of evaluation related to multicriteria preference elicitation provides users with more relevant detailed recommendations [2]. However, the implementation of such a model requires a knowledge base where the items are evaluated *w.r.t* a set of criteria. This constraint imposed by the model is very heavy for the user.

In [4], an unsupervised multicriteria opinion mining method is proposed. It allows users to free themselves from constraining partial evaluations *w.r.t* each criterion: users simply submit their critics in natural language to express their opinions, and system analyzes them automatically. It first dissects the critics according to the evaluation criteria (thematic segmentation) before calculating the polarity or opinions of each of the extracts resulting from the segmentation step (opinion-mining/sentiment analysis).

Combining this method with an interactive multicriteria decision support system makes possible to have a highly automated system for recommendation purposes. However, the automated assignment only provides imprecise scores related to items that are modeled by intervals in an adequate multicriteria analysis process. Our possibility theory based approach then manages multiple imprecise assessments derived from sentiment analysis on each evaluation criterion (intervals fusion) and then aggregate them on all criteria. Finally, we try to match a user and an adequate specialized magazine in the domain of concern (movies in our application) that will provide the most suitable personalized recommendation to the user.

Section 2 summarizes opinion mining approach to extract Internet user's critics and compute opinion scores on a set of criteria. Section 3 explains how to merge these imprecise opinion scores for each criterion and introduces the notion of matching of a distribution with data available on a criterion. Section 4 describes how to deduce the multicriteria model used by a specialized magazine to assess items. Section 5 shows how these approaches can be combined to address the multicriteria recommendation problem in the case study of movie recommendations.

## 2 Opinion-Mining Process

On the basis of statistical methods, the Opinion-mining approach of [4] allows us to build a lexicon of opinion descriptors for a given thematic. This lexicon is used to automatically extract the polarity of text segments that are related to the criterion. Two stages are distinguished in this multicriteria evaluation: - firstly, the segments of text related to each of the evaluation criteria are extracted with the *Synopsis* approach described in [5]. The text is first segmented into criteria. Then, for each criterion, the polarities of the segments that have been identified

by the *Synopsis* approach are computed. This is opinion-mining or sentiment analysis step.

### 2.1 Text Segmentation by Criterion

The Synopsis approach is used to identify the text extracts that refer to the adopted criteria and consists of 3 steps:

1. automatic construction of a training corpus used to learn characteristic words (or groups of words), called *descriptors*, for a criterion of interest
2. Automatic learning of these *descriptors* and construction of a lexicon associated with the criterion
3. Text segmentation using the lexicon associated with the criterion

The approach uses a set of *seed words*. On the one hand, they serve to semantically characterize the criterion of concern, and on the other hand to initiate the learning of *descriptors* of the criterion [5]. The training corpus is built automatically. The hypothesis of the learning is based on the fact that the more frequently a descriptor is found in the neighborhood of a *seed word* of the criterion (counting on sliding window), the greater the membership function of this descriptor to the lexical scope of the criterion.

Synopsis builds automatically its training corpus for each criterion by using web documents. This corpus is used to automatically build a lexicon of descriptors for each of these criteria. Each lexicon is then used by the segmentation process to automatically extracts parts of text dealing with this criterion. Synopsis will identify several segmentation for each criterion depending of granularity levels. Those granularities can be associated with user's level of expertise. The pertinent levels are automatically identified by Synopsis (see [4] and [5] for more details).

### 2.2 Opinion Analysis

The opinions extraction approach is an adaptation of the *Synopsis* approach to the extraction of opinions both for the automatic building of the training corpus and for the descriptors learning phases. Seed words become opinion seed words [4]. Two sets of words of opinion are initially distinguished the positive ones: $P = \{good, nice, excellent, positive, fortunate, correct, superior\}$; and the negative ones: $N = \{bad, nasty, poor, negative, unfortunate, wrong, inferior\}$.

Assuming a document that contains at least one word of P (resp. N) and none of N (resp. P) conveys a positive opinion (resp. negative), a set of documents associated with a seed word is built with a search engine on the web as in *Synopsis* (request for seed word "good" in the movie domains: movie + good - bad - nasty - negative - poor - unfortunate - wrong - inferior). Opinion descriptors are limited to adjectives and adjectival groups [4]. Statistical techniques of filtering with sliding windows of *Synopsis* are adapted to count opinion descriptors occurences.

Finally, an opinion score for a criterion is provided as a weighted sum of the text segments' membership related to the criteria and the positive/negative

descriptors. Thus, since the extraction of criteria segments depends on the level of expected precision (*i.e.* level of expertise), the score of the opinion text related to the criterion is also affected by the discrimination threshold. For each threshold the segmentation algorithm of [4] generates the corresponding text segmentation. Then the opinion scores of the text can be computed for any user's expertise level. There exists a lower and upper bound for this score. Accordingly, the opinion score is an imprecise entity which is then represented as an interval whose extremities are these lower and upper bounds.

## 3 Intervals Merging

The imprecision involved here concerns the subjectivity of a critic in the evaluation process of a text. This "subjectivity" is technically related to the imprecision of extraction. It is however "homogenized" by the automatic processing of segmentation. Then, evaluation is also uncertain because of the multiplicity of automatically collected opinions. Belief theory [6], [7] provides an appropriate framework to summarize these opinions and ease their representation and manipulation in the recommendation process while respecting their imprecision and uncertainty. Possibility distributions [8] are good approximations of belief functions. Thus they will be used to represent assessments. Also, possibility functions are appealing from an interpretation point of view in collecting confidence intervals as well as a computational point of view.

### 3.1 Possibility Theory

Let $\Omega$ represent a universal set of elements $\omega$ under consideration that is assumed to be finite and let $2^\Omega$ represent the power set of $\Omega$. A possibility distribution $\pi$ is a normalized function $\pi : \Omega \longrightarrow [0,1]$ (i.e. $\exists\omega \in \Omega$, such that $\pi(\omega) = 1$). From $\pi$, possibility and necessity measures are respectively defined for all subsets $A \in 2^\Omega$: $\Pi(A) = \sup_{\omega \in A} \pi(\omega)$ and $N(A) = 1 - \Pi(A^c)$. $\Pi(A)$ quantifies to what extent the event $A$ is plausible while $N(A)$ quantifies the certainty of $A$. An $\alpha$-cut of possibility distribution $\pi$ is the classical subset: $E_\alpha = \{\omega \in \Omega : \pi(\omega) \geq \alpha\}$, $\alpha \in ]0,1]$. When a distribution has a trapezoidal form, it is classically represented by its vertices abcd.

### 3.2 Evidence Theory

The evidence theory shall now be formulated by the basic belief assignment (*bba*) $m$ defined from $2^\Omega$ to $[0,1]$, such that: $\sum_{\{A \subseteq \Omega\}} m(A) = 1$ and $m(\emptyset) = 0$. Elements $E$ of $2^\Omega$ such that $m(E) > 0$ are called focal elements and their set is denoted $\mathtt{F}$. The *bba* $m$ can be represented by two measures: the belief function $Bel(A) = \sum_{\{E \in \mathtt{F}/A \supseteq E\}} m(A)$, $A \subseteq \Omega$ and the plausibility function $Pl(A) = \sum_{\{E \in \mathtt{F}/A \cap E \neq \emptyset\}} m(A)$, $A \subseteq \Omega$. When focal elements are imprecise, the probability of any event $A \subseteq \Omega$, denoted $Pr(A)$, is imprecise and $Bel(A)$ and

$Pl(A)$ represent respectively the lower and upper probabilities of event $A$, *i.e.* $Pr(A) \in [Bel(A), Pl(A)]$. Two well-known extreme cases of belief and plausibility measures are probability measures and possibility measures [8].

### 3.3 Building Possibility Distributions From a Set of Intervals

Let consider a set of distinct intervals $\{I_j, j = 1, nbi\}$ as the focal elements and the probability of occurrence of interval $I_j$ as the *bba* $m(I_j)$ assigned to this interval. When intervals are nested, *i.e.* $I_1 \subset I_2 \subset ... \subset I_{nbi}$, a possibility distribution $\pi$ may be built from plausibility measure, as proposed in [9]: $\forall \omega \in \Omega$, $\pi(\omega) = Pl(\{\omega\}) = \sum_{j=1,nbi} m(I_j).\mathbf{1}_{I_j}(\omega)$. When intervals are consistent, *i.e.* $\bigcap_{j=1,nbi} I_j = I \neq \emptyset$ (all experts share at least one value), but not nested, two possibility distributions $\pi_1$ and $\pi_2$ are built: First, we consider the *bba* $m_1(I_j)$ for focal elements $\{I_j, j = 1, nbi\}$. Thus $\forall \omega \in \Omega$, $\pi_1(\omega) = \sum_{j=1,nbi} m_1(I_j).\mathbf{1}_{I_j}(\omega)$. Second, $r$ nested focal elements $\{E_s, s = 1, r\}$ are obtained from original data from the $\alpha$-cuts of $\pi_1$: $E_1 = I$ and $E_s = E_{s-1} \cup E_{\alpha_s}(\pi_1)$ $(s = 2, r)$. The new bba $m_2$ assigned to intervals $E_s$ are computed as proposed in [9]: $m_2(E_s) = \sum_{\{I_j \ related \ to \ E_s\}} m_1(I_j)$ (each assessments $I_j$ being *related* in a unique way to the smallest $E_s$ containing it). Then a possibility distribution $\pi_2$ can be defined as: $\forall \omega \in \Omega$, $\pi_2(\omega) = \sum_{s=1,r} m_2(E_s).\mathbf{1}_{E_s}(\omega)$. Membership functions $\pi_1$ and $\pi_2$ are mono modal possibility distributions since $\bigcap_{j=1,nbi} I_j = I \neq \emptyset$ holds. Furthermore, they are the best possibilistic lower and upper approximations (in the sense of inclusion) of assessment sets $\{I_j, j = 1, nbi\}$ [9]. It can be seen easily that $\pi_1 \subseteq \pi_2$ (inclusion of fuzzy subsets) as $\forall \alpha \in ]0, 1]$, $E_{1,\alpha} \subseteq E_{2,\alpha}$.

In general however, experts' assessment might be neither precise nor consistent. The probability and possibility representations correspond respectively to extreme and ideal situations; unfortunately, critics may reveal contradictory assessments. This means that the consistency constraint may not be satisfied in practice, i.e. $\bigcap_{j=1,nbi} I_j = \emptyset$. To cope with this situation, groups of intervals, *maximal coherent subsets* (*MCS*), with a non-empty intersection are built from original intervals, which is equivalent to find subsets $K_\beta \subset \{1, ..., nbi\}$ with $\beta \in \{1, ..., g\}$ such that: $\bigcap_{j \in K_\beta} I_j \neq \emptyset$, with g being the number of subsets $K_\beta$ [10].

For each group $K_\beta$, lower and upper possibilistic distributions $\pi_1^\beta$ and $\pi_2^\beta$ are built (as in the previous case when elements are consistent). Let possibility distribution $\pi_1$ (resp. $\pi_2$) be the union (denoted $\tilde{\bigcup}$) of possibility distributions $\pi_1^\beta$ (resp. $\pi_2^\beta$):

$$\pi_1 = \tilde{\bigcup_{\beta=1,g}} \pi_1^\beta \ (resp. \ \pi_2 = \tilde{\bigcup_{\beta=1,g}} \pi_2^\beta) \tag{1}$$

then $\pi_1$ and $\pi_2$ are the multi-modal ($g$ modes) possibilistic lower and upper approximations of original intervals.

Reasoning with the lower distribution (resp. upper distribution) might correspond to a severe risk aversion position relative to the probability of information (resp. a flexible risk acceptance position). To maintain the richness of information provided by critics, the best way is to keep both distributions.

### 3.4 Matching between Distribution and a Set of Intervals

Let consider possibility distributions $\pi_1$ and $\pi_2$ the possibilistic approximations of intervals $\{I_j, j = 1, nbi\}$. Let denote for $\pi_1$ (resp. $\pi_2$) $(N_1, \Pi_1)$ (resp. $(N_2, \Pi_2)$) the associated possibility and necessity measures. Then $\pi_1$ and $\pi_2$ are respectively the *greatest* and *smallest* fuzzy subsets such that [9]: $\forall A \subseteq \Omega, [N_1(A), \Pi_1(A)] \subseteq [Bel(A), Pl(A)] \subseteq [N_2(A), \Pi_2(A)]$. Let still consider two possibility distributions $\pi$ and $\pi^*$ defined on $\Omega$. Two definitions are introduced:

**Definition 1.** *We define the degree of inclusion of $\pi$ in $\pi^*$ as:*

$$incl(\pi, \pi^*) = (\int_\Omega (\pi^* \wedge \pi))/\int_\Omega \pi \tag{2}$$

**Definition 2.** *We define the degree of matching of $\pi$ to data $\{I_j, j = 1, nbi\}$ as:*

$$match(\pi, \{I_j\}) = [incl(\pi_1, \pi) + incl(\pi, \pi_2)]/2 \tag{3}$$

*$\{I_j\}$ is said to match $\pi$ better than $\pi^*$ if: $match(\pi^*, \{I_j\}) < match(\pi, \{I_j\})$. We also use the notation $match(\pi, (\pi_1, \pi_2))$ instead of $match(\pi, \{I_j\})$ when possible.*

*Remark 1.* Definition 2 leads, for particular cases of $\pi$, to:

- $match(\pi, \{I_j\}) \in [0, 1]$.
- If $\pi_1 \subseteq \pi \subseteq \pi_2$ (inclusion of fuzzy subsets), then $match(\pi, \{I_j\}) = 1$.
- If $\pi_2 \subset \pi$, then $match(\pi, \{I_j\}) = [1 + (\int_\Omega \pi_2 / \int_\Omega \pi)]/2$.
- If $\pi \subset \pi_1$, then $match(\pi, \{I_j\}) = [(\int_\Omega \pi / \int_\Omega \pi_1) + 1]/2$.
- If $\pi_2 \cap \pi = \emptyset$, then $match(\pi, \{I_j\}) = 0$.

The idea behind the *matching* is to consider that the distribution $\pi$ (with $N$ and $\Pi$ its associated possibility and necessity measures) which guarantees $[N_1(A), \Pi_1(A)] \subseteq [N(A), \Pi(A)] \subseteq [N_2(A), \Pi_2(A)]$ for a large number of subset $A \subseteq \Omega$, *better* matches data $\{I_j, j = 1, nbi\}$. This approach is similar to the one in [11] which used in fuzzy pattern matching.

## 4 Identification of Preferences Model

Aggregation models make the capture of the notion of priorities in the decision-maker's strategy possible, and simplify the comparison of any two alternatives described through their elementary evaluation. The most commonly used operator to express decision maker preferences is the weighted average mean denoted

here $WAM_\omega$. It allows giving non-symmetrical roles to criteria through a vector of relative weights $\omega = (\omega_1, ..., \omega_n) \in [0,1]^n$.

The specialized press generally provides a simple overall assessment of items using evaluation scales such as: number of stars, number of bars, etc.. This score is accompanied by a more or less detailed critic in natural language which is supposed to make explicit the assessment of the journal. However it is often difficult for users to understand the exact reasons that would justify the imprecise overall score the item received. We model these imprecise scores by possibility distributions $\widetilde{\pi}^k$ on $[0, 20]$ where k is an item index. To make clearer assessments allocated by a magazine to an item, we try to identify the strategy, the priorities that characterize this journal: what are the criteria that differentiate the values conveyed by this journal? As soon as a user knows which critics are of interest in the assessments of a specialized magazine, he can then choose the journal that matches the best his priorities and choose a film recommended by this magazine.

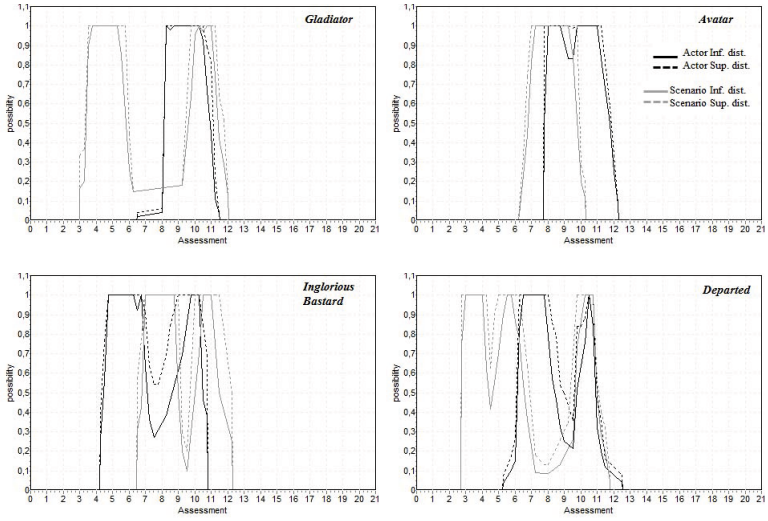
Many Internet users contribute to collaborative recommender systems but provide their opinion in natural language because they are not familiar with assessments and all the less with multi criteria assessments. Our segmentation and sentiment analysis system automatically collects all these criteria and, for each critic $c$ that deals with an identified criterion $i$, it extracts an imprecise score (depending of the analysis granularity, *i.e.* the expected level of expertise) which is modeled as an interval as explained in section 2.2. According to the merging method presented in section 3.3 we compute two possibility distributions $\pi^k_{i,1}$ (inferior) and $\pi^k_{i,2}$ (superior) for the set of automatically collected imprecise scores for each item $k$ and each criterion $i$. Let us note $\overline{\pi_{\alpha,\omega}}^k = WAM_\omega(\pi^k_{\alpha,1}, ..., \pi^k_{\alpha,n})$ with $\alpha \in \{1, 2\}$. The next step is to identify the weights of the distribution that best match the magazine's overall assessments. In other words, we search for weights $\omega$ such that $\widetilde{\pi}^k$ and $(\overline{\pi_{1,\omega}}^k, \overline{\pi_{2,\omega}}^k)$ match as well as possible for a learning set of items (items). Mathematically, a possible answer is based on our function *match* defined in section 3.4 such as:

$$\omega^* = Arg \max_{\omega \in [0,1]^n} \sum_{k \in items} match(\widetilde{\pi}^k, (\overline{\pi_{1,\omega}}^k, \overline{\pi_{2,\omega}}^k)) \tag{4}$$

## 5 Case Study

The software prototype that supports our recommendation system is based on a combination of an Internet user with a specialized journal that bears the same priorities or values as him. The case study presented in this section is based on movie recommendations. This prototype uses the multicriteria opinion extraction module in section 2, as well as, a base of movie critics written in natural language from the famous film critics site IMDB. Critics provided by IMDB (about 3000 critics per movie) provide enough information to get a representative picture of the diversity of opinions about this movie. Each film critic has been evaluated by our multicriteria opinions extraction system. To illustrate the method and simplify the presentation only two criteria are considered in this toy case study:

**Table 1.** Inferior and superior possibility distributions




*actor* and *scenario*. With the merger process of Section 3.3, we obtain inferior and superior distributions for 20 films in this case study. Table 1 shows the results respectively for movies: Gladiator, Avatar, Inglorious Bastard and Departed.

As we can see in table 1 multi modality is present in all distributions. It is due to the divergence found in critics' opinions: users are rarely unanimous on a movie review.

We selected 39 journals in this application. For each movies of the base, some journals provide overall assessments in the form of a number of stars (transformed into trapezoidal distributions $\widetilde{\pi}^k$) (see table 5). The weight distribution that characterize the best the journal evaluation strategy is calculated through equation 4 for each journal: *e.g.* the weight distribution that explains at best the scores $\widetilde{\pi}^k$ assigned to the 39 movies by the journal in the learning database. Fig. 1 shows that there are large differences between assessment strategy of journals (*e.g.* for Cahiers du Cinema, the weights are 0.6 for *actor* and 0.4 for *scenario* while conversely they are 0.22 and 0.78 for Charlie Hebdo). Some journals attach no importance to actors and some others to scenario.

Note that for a reliable and relevant recommendation, our model should integrate more criteria in the assessment. Finally, Fig. 1 provides to the Internet user how important criteria are considered in the 39 journals assessment strategy. He

**Table 2.** Transformation of stars to trapezoidal distributions

| numbers of stars | */2 | * | *(*/2) | ... | **** |
|---|---|---|---|---|---|
| abcd trapezium | {0,2.5,5,7.5} | {2.5,5,7.5,10} | {5,7.5,10,12.5} | ... | {12.5,15,17.5,20} |

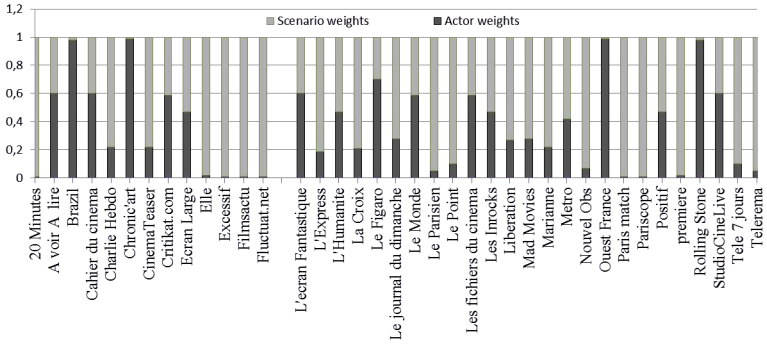



**Fig. 1.** Assessment strategy of movies specialized journals

may then simply choose the journal that conveys the values that are closest to his mood of the moment and consult the hit list of this journal for a personalized recommendation. No model of user preference is need to be identified which is generally a thorny and forbidding task in collaborative RS systems. The user identifies himself the journal that suits him as *best*. Note that this simple principle allows the user to change his "preferences system" depending on his mood every time he goes to the cinema!

## 6 Conclusion

The automated extraction of critics related to a set of criteria, the imprecise assessment process based on our sentiment analysis and our fuzzy multicriteria analysis allows the development of highly automated recommender systems of type "multicriteria preference elicitation from evaluations"[2] free of the most constraining tasks of this type of collaborative systems. This is an important step because until now the need to manually assess a large number of documents according to several criteria represented a major obstacle to the implementation of such systems. The process we propose establishes a cognitive automation that can be easily deployed into Web applications.

## References

1. Imoussaten, A., Duthil, B., Trousset, F., Montmain, J.: Possibilistic interpretation of a sentiment analysis for a web recommender system. In: LFA 2013, Reims (2013)
2. Adomavicius, G., Manouselis, N., Kwon, Y.: Multi-criteria recommender systems. In: Recommender Systems Handbook, pp. 769–803. Springer (2011)
3. Manouselis, N., Costopoulou, C.: Analysis and classification of multi-criteria recommender systems. World Wide Web 10(4), 415–441 (2007)
4. Duthil, B., Trousset, F., Dray, G., Montmain, J., Poncelet, P.: Opinion extraction applied to criteria. In: Liddle, S.W., Schewe, K.-D., Tjoa, A.M., Zhou, X. (eds.) DEXA 2012, Part II. LNCS, vol. 7447, pp. 489–496. Springer, Heidelberg (2012)

5. Duthil, B., Trousset, F., Roche, M., Dray, G., Plantié, M., Montmain, J., Poncelet, P.: Towards an automatic characterization of criteria. In: Hameurlain, A., Liddle, S.W., Schewe, K.-D., Zhou, X. (eds.) DEXA 2011, Part I. LNCS, vol. 6860, pp. 457–465. Springer, Heidelberg (2011)
6. Dempster, A.P.: Upper and lower probabilities induced by multivalued mapping. Annals of Mathematical Statistics 38 (1967)
7. Shafer, G.: A mathematical theory of evidence. Princeton University Press (1976)
8. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
9. Dubois, D., Prade, H.: Fuzzy sets and statistical data. European Journal of Operational Research 25 (1986)
10. Imoussaten, A., Mauris, G., Montmain, J.: A multicriteria decision support system using a possibility representation for managing inconsistent assessments of experts involved in emergency situations. IJIS 29(1), 50–83 (2014)
11. Dubois, D., Prade, H., Testemale, C.: Weighted fuzzy pattern matching. Fuzzy Sets and Systems 28(3), 313–331 (1988)

# Suggesting Recommendations Using Pythagorean Fuzzy Sets illustrated Using Netflix Movie Data

Marek Z. Reformat[1] and Ronald R. Yager[2]

[1] University of Alberta, Edmonton, Canada
Marek.Reformat@ualberta.ca
[2] Iona College, NY, USA
yager@panix.com

**Abstract.** The web can be perceived as a huge repository of items, and users' activities can be seen as processes of searching for items of interest. Recommender systems try to estimate what items users may like based on similarities between users, their activities, or on explicitly specified preferences. Users do not have any influence on item selection processes.

In this paper we propose a novel collaborative-based recommender system that provides a user with the ability to control a process of constructing a list of suggested items. This control is accomplished via explicit requirements regarding rigorousness of identifying users who become a reference base for generating suggestions. Additionally, we propose a new way of ranking items rated by multiple users. The approach is based on Pythagorean fuzzy sets and takes into account not only assigned rates but also their number. The proposed approach is used to generate lists of recommended movies from the Netflix competition database.

## 1 Introduction

The most common activity of users on the web is looking for items of interest or potential interest. The two well-know tools supporting users in these activities are search engines and recommender systems [1][2].

In recommender systems, items presented to the user are selected using variety of techniques. The two most popular ones are collaborative and knowledge-based approaches [3][4]. In the former approach, items are being suggested based on the items seen, bought or simply rated by other users who seen, bought or rated, in the past, the same items as the user who is looking for recommendations. In the knowledge-based approach, the user provides her preferences and constrains, and recommended items match these requirements to the highest possible degree. An approach using a description of the items and the user's interests, called content-based, is gaining popularity [x][y].

The most important task of recommender systems it to provide the users with items that represent a best possible match to their interests, and in many cases this fits the users' needs. However, if the user wants to be provided with different items, the performance of these systems is not very satisfying. Any deviation from "perfectly"

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 546–556, 2014.

matched items seems to be accidental. The user does not have any control over these "extensions" or "additions". The process of recommendation of items is controlled by history of the user's activities, or directly via the user's preferences.

Another vital issue for recommender systems is processing multiple ratings associated with a single item. It is always a question how to balance the ratings themselves with their number, and how to combine them to obtain a single score identifying goodness of an item.

In the paper we propose application of fuzziness to address the issue of the user's control of a selection process in a collaborative recommender system, as well as to determine a single score describing an item based on multiple ratings. The user controls a selection process via explicitly shaping composition of a group of users whose items and ratings constitute a reference base for building a list of suggested items. The process of calculating a single score is based on Pythagorean fuzzy sets proposed in [7][8].

The outline of the paper is as follows. Section 2 contains an overview of the whole process of building a list of suggestions. The user-controlled process of building a group of users is fully described in Section 3, while details of creating a ranking of items are included in Section 4. Section 5 contains description of a real-world example using the Netflix competition database. The paper is summed up with a conclusion section.

## 2 Fuzzy-Controlled Suggestions: Overview

When a list of suggested items is provided to the user, it would be good to see some variety of items shown to the user. This would resemble a true browsing process – the user "roughly" knows what she wants and tends to wander around looking for things that could be of interest for her. Of course, the user is doing it with all her "background" knowledge regarding what she likes, but at the same time, she lets herself to be adventurous. In order to mimic such behaviour, we propose the following procedure that ensures a more human-like experience in finding variety of items that could have potential interest to the user.

The first step is to determine a group of users that have evaluated the same items (for example movies) in a similar way as the user who looks for suggestions. The similarity is controlled by the user, and reflects her way of determining a level of matching between her evaluations and evaluations of other individuals who experienced (watched) the same items (movies) as she. So far, we envision three possible ways of identifying similarity of items (movie) evaluations: a) strict – evaluations have to match to the highest possible degree, b) strict on high – positive evaluations have to match to the highest possible degree, however for the negative evaluations the user is forgiving, in other words, the user wants to see a list of suggestions created based on users with who agree with her in the case of the positive evaluations and could have a different opinion in the case of negative evaluations, and c) strict on low – the opposite to the above one, i.e., the user wants to see a list of suggestions created based on users with whom she agrees regarding the negative evaluations, and relaxes her requirements for matching positive evaluations. This process will lead a list of users with similar – up to the user's requirements – preferences regarding items (movies). Each user's preference pattern is determined by aggregation of all matching scores across of all items (movies) overlapping between the user and a potential member of a group.

The **second step** is to identify a list of items that have not been experienced by the user who looks for suggestions, but common to all the users from the group. It is the

simplest step – it involves finding a common set of items among all items experienced by the members of the group identified in the first step.

The **third step** is to rank all selected items. The ranking of items has to take into consideration two components: quantitative – a number of ratings for each item, and qualitative – a distribution of positive and negative ratings.

# 3 Building User Groups with Different Levels of Strictness

## 3.1 General Overview

In order to provide any suggestions we need to have a pool of users who somehow resembles or have similar interests to the user who is looking for suggestions. Selection of such a group can be done in multiple ways, and in majority of situations such as process happens without any influence of the user. The approach we propose here provides the user with a way of indicting how "close" members of created group should be when compared with the user. The user governs this process via providing linguistic terms identifying strictness of the comparison between items rated by the user and items rated by members of a group being created.

## 3.2 Comparison of Ratings

Any items rated by two users can be compared using different approaches. The simplest one would be to check if both ratings are the same. In such a case, we would obtain a binary result – a perfect match or total mismatch. Quite often, such a comparison is useful and it leads to finding a person who is very much like the user, kind of "a mirror image of the user". However, in the situation the user would like to "expend" her set of items and "go beyond" its own comfort zone, i.e., find something more diverse – a different type of comparison is needed. The approach presented here addresses such a need.

A proposed approach for comparison of ratings uses linguistic terms to control flexibility of this process, i.e., how differences between ratings should be treated. There are a number of different rating schemas, but without loss of generality we assume for the rest of the paper that the ratings are in the range from 1 to 5, where 1 represents "do not like" and 5 "like very much", with 3 indicating "neutral". To make the comparison of ratings controllable we identify three terms describing strictness of the comparison process:

- *two-side bounded* evaluation: the comparison is very strict, both ratings – positive and negative – have to be matched to the highest degree;
- *positive-side bounded* evaluation: the comparison is strict for positive evaluations, and relaxed for negative evaluations, in other words the user is okay when her negative ratings are not considered very "seriously", but positive ratings have to be matched to the highest degree;
- *negative-side bounded* evaluation: the comparison is opposite to the above one – positive ratings do not have to match perfectly, but the negative ones have to be respected.

Let us assume we have two items to compare: an item to which the comparison is done – we called it a *reference_item*, and an item being compared – we called it an *other_item*. The comparison procedure is as follows: 1) the rating $R_{ref}$ of the

*reference_item* is subtracted from the rating $R_{other}$ of the *other_item*, 2) the obtained difference is modified by an appropriate mapping representing the user's evaluation requirement, and the resulting value is a level of compatibility of two ratings.

An implementation of terms controlling the evaluation process is done using fuzzy sets. Let us define a universe of discourse **D** that is a range of differences between ratings. A fuzzy set is a mapping

$$\mu_{LTerm}: D \rightarrow [0,1],$$

and the compatibility of ratings is

$$RateComp = \mu_{LTerm}(R_{other} - R_{ref})$$

where $\mu_{LTerm}$ is a fuzzy set associated with an appropriate linguistic term. Three possible sets representing the described above terms are shown in Fig 1.

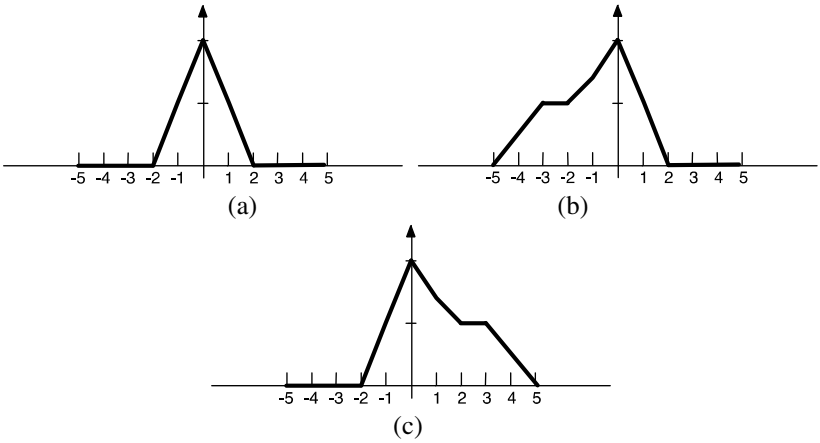



**Fig. 1.** Examples of fuzzy sets representing three linguistic terms: *two-side bounded* (a), *positive-side bounded* (b), *negative-side bounded* (c)

Fig 1(a) represents a fuzzy set for the *two-side bounded* evolution. It allows for a small deviation of $R_{other}$ from $R_{ref}$: the perfect match gives the score of 1.0, while a difference of 1.0 or -1.0 leads to the score of 0.5. Any other difference gives the score of 0.0. A fuzzy set for the *negative-side bounded* evaluation, Fig 1(b), gives the score of 1.0 for the perfect match. However, it "penalizes" the situation when the difference $R_{other}$-$R_{ref}$ is positive, what ensures that the user's low ratings are well matched. On the other hand, when the difference $R_{other}$-$R_{ref}$ is negative, i.e., when $R_{other}$ is smaller than $R_{ref}$, the evaluation is more forgiving – there is no requirement for a good match for the user's high ratings. A fuzzy set presented in Fig 1(c) for the *positive-side bounded* evaluation reflects the inverse behaviour to the *negative-side bounded* case.

### 3.3 Group Generation

A group is built in a process of selecting users who have a similar pattern of rating items to the user who looks for suggestions. Using different comparison fuzzy sets we obtain different groups of users. Each of these groups contains users who match the

user (our reference user) differently. This effect is desirable because this means that each group contains a slightly (not totally) different set of users. And a different set of users leads to a different set of items – therefore, suggestions provided by each group can be a bit different.

In a nutshell, a process of creating a group of users can be described via the following algorithm:

**INPUT:** $M_{ref}$ movies rated by the user $U_{ref}$
$UL_{all}$ a list of users
$\{M_i\}$ a set of movies rated by each user ($U_i$) from the list $UL_{all}$
**OUTPUT:** G a list of users constituting a group

1 Select Users who ranked the Same Movies as $U_{ref}$, i.e., $M_i$ contains $M_{ref}$
2 Create a list of Selected Users: $UL_{other}$
3 **for** each user $U_i$ from $UL_{other}$
4 **for** each movie from $M_{ref}$
5 Compare Ratings: calculate score and store it
6 **rof**
7 Create Rating Compatibility measure: aggregate scores
8 **rof**
9 Sort Users from $UL_{other}$ based on Rating Compatibility measures
10 Create group G: select top 25 percent from sorted $UL_{other}$

Another human-like aspect of creating a group is a linguistic-based control of an aggregation process (line 7). In the proposed approach we use linguistic-based OWA [9]. Any linguistic quantifiers can be applied to calculate a rating compatibility measure.

### 3.4 Example

To illustrate the process let us take a look at a very simple situation where we have the $U_{ref}$ and another user – $U_i$, and both of them ranked the same five items. Their ratings are in Table 1 column 1 and 2, respectively.

**Table 1.** Illustration of a proposed method for composing a group of users

| $U_{ref}$ | $U_i$ | difference $R_i$-$R_{ref}$ | two-side | positive-side | negative-side |
|---|---|---|---|---|---|
| ratings | | | Rating Compatibility | | |
| 5 | 1 | -4 | 0 | 0.25 | 0 |
| 4 | 2 | -2 | 0 | 0.5 | 0 |
| 3 | 3 | 0 | 1 | 1 | 1 |
| 2 | 4 | 2 | 0 | 0 | 0.25 |
| 1 | 5 | 4 | 0 | 0 | 0.5 |

The difference between ratings is shown in column 3. This different is use as the input to three different linguistic terms introduced in the previous section: *two-side bounded* – the obtained values are in column 4; *positive-side bounded* – column 5;

and *negative-side bounded* – column 6. It can be easily observed that the *two-side bounded* term is very strict, the *positive-side bounded* term "forgives" when there is a mismatch for items ranked high by $U_{ref}$, and the *negative-side bounded* term "forgives" when the mismatch happens for items which the $U_{ref}$ ranked low.

The application of OWA with the linguistic quantifier MOST leads to a small compatibility score of 0.04 in the case of the *two-side bounded* evaluation. For the *negative-* and *positive-side bounded* evaluation the score is the same 0.15.

# 4 Ranking of Items with Multiple Ratings

## 4.1 Overview

The group generation process ends up with a group of users who rank items in a way that resembles the user's way, i.e., it could "tightly" resemble the user's rankings for both positive and negative rankings, or just one of them. At the same time many of these users ranked items that have not been experienced/ranked by the user yet, and a set of these items is a starting point for building a suggestion list.

In general multiple users from the group can rank, in the scale from 1 (worst) to 5 (best), a single item. It means that a single item can have multiple rankings. The proposed approach is taking into consideration two aspects in order to determine a single score representing a degree of attractiveness to the user. These two aspects are: overall combined ratings provided by the users from the group; and a number of users who provided these ratings. Such calculated degree of attractiveness is used to sort item and provide a list of suggestions to the user. In the proposed method we use Pythagorean fuzzy sets to identify the final ranking of suggestions.

## 4.2 Pythagorean Fuzzy Sets

The Pythagorean fuzzy sets (PFS) have been proposed by Yager in [7][8]. It is a class of new non-standard fuzzy sets. A membership grade of PFS can be expressed as: $r(x)$ called the strength of commitment, and $d(x)$ called the direction of commitment, for each x from the domain X. Both of them are in the range from 0.0 to 1.0. Here, $r(x)$ and $d(x)$ are linked with a pair of membership grades $A_Y(x)$ and $A_N(x)$. $A_Y(x)$ represents support for membership of $x$ in $A$, while $A_N(x)$ represents support against membership of $x$ in $A$. The relations between all these measures are:

$$A_Y^2(x) + A_N^2(x) = r^2(x)$$

$$A_Y(x) = r(x) * Cos(\theta)$$

$$A_N(x) = r(x) * Sin(\theta)$$

where

$$\theta = \arccos\left(\frac{A_Y(x)}{r(x)}\right)$$

and

$$\theta = (1 - d(x)) * \frac{\pi}{2}$$

The value of $r(x)$ allows for some lack of commitment. Its value is in the unit interval and the larger $r(x)$ is the stronger the commitment is, and the less the

uncertainty it represented. The value of direction of commitment $d(x)$, on the other hand, can provide an interesting insight into relations between $A_N(x)$ and $A_Y(x)$. For $\theta = 0$, i.e., when there is not $A_N(x)$ part, the value of $d(x) = 1$ and this means that there is no "negative" comments and the commitment direction is fully positive. When both $A_Y(x)$ and $A_N(x)$ are equal, the angle $\theta = \pi/4$ and the value of $d(x) = 0.5$ indicating a natural direction of commitment. The other boundary condition is when only $A_N(x)$ is present and there is no $A_Y(x)$ – in this case $d(x) = 0.0$ representing a lack of positive direction of commitment . As you can see the value of $d(x)$ is "detached" from the number of comments, it only depends on the ratio of comments. It is a very interesting feature of PFS that is used in the proposed approach.

### 4.3 Pythagorean-Based Ranking

A process of building a ranking starts with determining a score, called degree of attractiveness, for a single item. The process comprises a number of steps.

The first step is to build PFS for a single item. In order to do this we perform a mapping from the users' ratings into $A_Y(x)$ and $A_N(x)$ values. It happens in the following way:

- ratings 5 and 4 are mapped into 1.0 and 0.5, and the sum of them is assigned to $A_Y(x)$
- ratings 1 and 2 are mapped into 1.0 and 0.5, and the sum of them is assigned to $A_N(x)$
- rating 3 is natural and is not mapped at all

Based on obtained values of $A_Y(x)$ and $A_N(x)$, we calculate values $r(x)$ – strength of commitment, and $d(x)$ – direction of commitment. This is performed for all movies. We find $r_{max}$ and normalize all $r(x)$. Now, the maximum $r(x)$ is 1.0, and all other ones are in the range from 0.0 to 1.0.

The next step is to calculate a score – an attractiveness value. And this is done using a single transformation of $d(x)$ and $r(x)$. If we represent a PFS in polar coordinates (Fig 2), we can think of an area bounded by the axe $A_N(x)$, $r(x)$, and a fragment of the circle circumference connecting $A_N(x)$ and the tip of the $r(x)$ – a thick line in Fig 2 (a). Such defined fragment of the circle represents simultaneously two things: a level of commitment (normalized $r_{norm}(x)$) and a direction of commitment – in this particular case more commitment means smaller angle $\theta$ and larger the circle fragment. The formula representing this relationship is:

$$score_1 = r_1 * \alpha_1 = r_1(\frac{\pi}{2} - \theta_1)$$

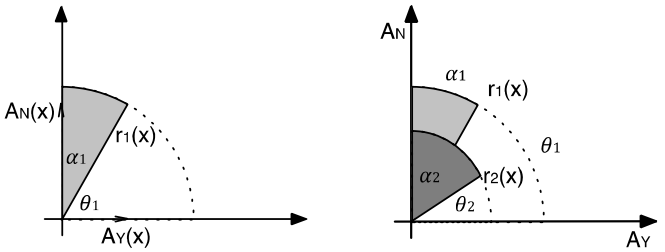
and knowing that

$$\theta_1 = (1 - d_1) * \frac{\pi}{2}$$

so

$$score_1 = r_1 d_1 \frac{\pi}{2}$$

As we can see, the attractiveness is simply a product of $r(x)$ (strength commitment) and $d(x)$ (direction of commitment), the $\pi/2$ is a simple constant that can be omitted for in the comparison process. If we calculate score for another PFS, Fig 2(b), we can compare both PFSs.




(a) (b)

**Fig. 2.** PFSs in polar coordinates: a single PFS (a), and two PFS with different r(x) and d(x) (b)

With such defined attractiveness of an item, ranking of items rated by the users for a determined group is a simple sorting process. As the result we obtain a list of suggestions.

### 4.4 Example

As in the case of building a group of users, also here we present a simple example that explains how five items with multiple ratings are compared, and eventually ranked. Let us assume the ratings as shown in Table 2, column 1.

**Table 2.** Illustration of a proposed method for ranking items with multiple ratings

| ratings | | $A_Y$ | $A_N$ | r | $r_{norm}$ | d | score | rank |
|---|---|---|---|---|---|---|---|---|
| Item-1 | 5, 4, 3, 2, 1 | 1.5 | 1.5 | 2.1213 | 0.6975 | 0.50 | 0.3486 | **3** |
| Item-2 | 5, 3, 1 | 1.0 | 1.0 | 1.4142 | 0.4650 | 0.50 | 0.2325 | **4** |
| Item-3 | 5, 4 | 1.5 | 0 | 1.5000 | 0.4932 | 1.00 | 0.4932 | **2** |
| Item-4 | 5, 4, 4, 4, 3, 2 | 2.5 | 0.5 | 2.5495 | 0.8383 | 0.87 | 0.7293 | **1** |
| Item-5 | 4, 3, 2, 2, 1, 1 | 0.5 | 3.0 | 3.0414 | 1.0000 | 0.11 | 0.1100 | **5** |

Table 2 contains all intermediate results, however we focus on the last four columns: $r_{norm}$ – normalized strength of commitment, d – direction of commitment, as well as score and rank. As we can see the strongest commitment has Item-5, and the weakest Item-2. In the case of direction of commitment – the most positive has Item-3 and Item-4, the most negative Item-5. Overall, the score – a product of $r_{norm}$ and d indicates Item-4 as the most suitable suggestion, with Item-3 as the second (even if it has only two ratings but both positive), and Item-1 as the third (quite a number of ratings but evenly split between positive and negative).

## 5 Case Study

### 5.1 Overview and Data Set Description

In order to illustrate the proposed method in a real-world scenario, we have applied it to building a list of suggested movies. For this purpose we use Netflix data that has

been used in Netflix competition [10]. The database of movies and ratings contains a total of 17770 movies and 480190 users.

We have created a user $U_{ref}$ who wants a suggestion list created from Netflix movies. We assume that the $U_{ref}$ have seen fifteen movies. A sample of these movies and the ratings assigned to them by the user are presented in Table 3.

**Table 3.** A sample of movies and their ratings

| **movie** | | **rating** |
|---|---|---|
| **title** | **ID** | |
| Star Trek: Voyager: Season 7 | 10141 | 5 |
| Star Trek: Insurrection | 12513 | 5 |
| Star Trek: The Next Generation: Season 1 | 10666 | 5 |
| The Sound of Music | 12074 | 2 |
| The Exorcist 2: The Heretic | 9387 | 2 |
| The Exorcist | 16793 | 1 |
| Night of the Living Dead | 9940 | 1 |

### 5.2 Group Building Process

The first step is to determine a group of users who "fit" the $U_{ref}$ requirements regarding compatibility of movie ratings. We used all three terms introduced in Section 3. As the result we obtained three groups of users:

- **group T-SB:** obtained with the *two-side bounded* term; it contains 2 users (their userIDs 1578801 and 647979), and a number of movies rated by the users is 4972;
- **group N-SB:** obtained with the *negative-side bounded* term: a number of users in the group is 6, and a number of different rated movies is 11034;
- **group P-SB:** obtained with the *positive-side bounded* term: the group is the largest, it includes 8 users, and a number of different movies to be selected from is 17540.

### 5.3 Generation of Recommendations

The movies associated with each of the groups, and not seen by the $U_{ref}$, have been used to create the suggestion lists. Due to the space limitations only top three movies, their scores and some statistics are shown:

- **group T-SB:** top movies are Akira Kurosawa's Dreams, The Good, and the Bad and the Ugly, Chariots of Fire; a total of 562 movies have a score above 0.5, including 235 with a perfect score of 1.0;
- **group N-SB:** top movies are The Lord of the Rings: The Fellowship of the Ring: Extended Edition, Lord of the Rings: The Return of the King, and Star Wars: Episode IV: A New Hope; in this case 421 movies obtained a score above 0.5, including 8 with 1.0;
- **group P-SB:** top movies are Lord of the Rings: The Return of the King, Lord of the Rings: The Fellowship of the Ring, and Lord of the Rings: The Two Towers; a total of 130 movies with a core above 0.5, only 3 with 1.0 score.

Please note that the recommendations generated based on the group T-SB contain the highest number of movies, and almost half of them have a perfect score. The fact that the $U_{ref}$ is very strict here shows that only two users from 480190 watched the same movies and rated them in the same way. Therefore, the suggestions do not have a huge "base" – only two users have rated each movie. This explains a high number of movies with a perfect score, and somehow diverse set of recommendations.

For the recommendations built using the other two groups, N-SB and P-SB, the situation is different: more users match the movie ratings made by $U_{ref}$, so the lists of suggestions seem to include movies that appears to be a good fit to $U_{ref}$ interests. There are a much smaller number of movies with the perfect score, but this can be explained by the fact that it is more difficult to find consensus among a larger number of users.

## 6 Conclusion

The users' anticipations regarding what can be found on the web exceeds initial expectations of finding items that represent a perfect match to what the users already know. On many occasions, the users want to be exposed to variety of items, items that go beyond their core interests. The recommendation systems try to provide the user with diversity of suggested items, but it is done without the users' input.

In the paper we introduce two fuzzy-based techniques that allow the user to control a degree of variation in recommended items. The user can be very strict and look only for items that fit her interests, or quite forgiving and open for related suggestions outside her interests. The first technique allows the user to control a process of identification a group of "reference users". Items rated by members of such a group are used to create an initial list of possible suggestions. The second technique is applied for generating a final ranking of items. It combines multiple ratings provided by members of the reference group, and generates a single score that reflects both number of ratings and their values.

The new methods have been used to generate lists of suggested movies using Netflix database.

## References

[1] Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 30, 107–117 (1998)

[2] Bobadilla, J., Ortega, F., Hernando, A., Gutierrez, A.: Recommender systems survey. Knowledge-Based Systems 46, 109–132 (2013)

[3] Terveen, L., Hill, W.: Beyond Recommender Systems: Helping People Help Each Other. In: HCI in the New Millennium, pp. 487–509 (2001)

[4] Burke, R.: Knowledge-Based Recommender Systems. In: Encyclopedia of Library and Information Systems (2000)

[5] Pazzani, M.J., Billsus, D.: Content-Based Recommendation Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007)

[6] Kim, B.M., Li, Q., Park, C.S., Kim, S.G., Kim, J.Y.: A new approach for combining content-based and collaborative filters. J. Intelligent Information Systems 27, 79–91 (2006)
[7] Yager, R.R.: Pythagorean Fuzzy Subsets. In: IFSA/NAFIPS Joint World Congress, pp. 57–61 (2013)
[8] Yager, R.R.: Pythagorean Membership Grades in Multi-Criteria Decision Making. IEEE Transactions on Fuzzy Systems (online August 212013)
[9] Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on Systems, Man and Cybernetics, 193–190 (1988)

# Evidential Communities for Complex Networks

Kuang Zhou[1,2], Arnaud Martin[2], and Quan Pan[1]

[1] School of Automation, Northwestern Polytechnical University,
Xi'an, Shaanxi 710072, P.R. China
[2] IRISA, University of Rennes 1, Rue E. Branly, 22300 Lannion, France
kzhoumath@163.com, Arnaud.Martin@univ-rennes1.fr, quanpan@nwpu.edu.cn

**Abstract.** Community detection is of great importance for understanding graph structure in social networks. The communities in real-world networks are often overlapped, *i.e.* some nodes may be a member of multiple clusters. How to uncover the overlapping communities/clusters in a complex network is a general problem in data mining of network data sets. In this paper, a novel algorithm to identify overlapping communities in complex networks by a combination of an evidential modularity function, a spectral mapping method and evidential $c$-means clustering is devised. Experimental results indicate that this detection approach can take advantage of the theory of belief functions, and preforms good both at detecting community structure and determining the appropriate number of clusters. Moreover, the credal partition obtained by the proposed method could give us a deeper insight into the graph structure.

**Keywords:** Evidential modularity, Evidential $c$-means, Overlapping communities, Credal partition.

## 1 Introduction

In order to have a better understanding of organizations and functions in the real networked system, the community structure, or the clustering in the graph is a primary feature that should be taken into consideration [3]. As a result, community detection, which can extract specific structures from complex networks, has attracted considerable attention crossing many areas from physics, biology, and economics to sociology [1], where systems are often represented as graphs.

Generally, a community in a network is a subgraph whose nodes are densely connected within itself but sparsely connected with the rest of the network [17]. Many of the community detection approaches are in the frame of probability theory, that is to say, one actor in the network can belong to only one community of the graph [9,4]. However, in real-world networks, each node can fully or partially belong to more than one associated community, and thus communities often overlap to some extent [11,15]. For instance, in collaboration networks, a researcher may be active in many areas but with different levels of commitment, and in social networks, an actor usually has connections to several social groups like family, friends, and colleagues. In biological networks, a node might have multiple functions [11].

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 557–566, 2014.

In the last decades, for identifying such clusters that are not necessarily disjoint, there is growing interest in overlapping community detection algorithms. Zhang et al. [17] devised a novel algorithm to identify overlapping communities in complex networks based on fuzzy $c$-means (FCM). Nepusz et al. [8] created an optimization algorithm for determining the optimal fuzzy membership degrees, and a new fuzzified variant of the modularity function is introduced to determine the number of communities. Havens et al. [5,6] discussed a new formulation of a fuzzy validity index and pointed out this modularity measure performs better compared with the existing ones.

As can be seen, most of methods for uncovering the overlapping community structure are based on the idea of fuzzy partition, which subsumes crisp partition, resulting in greater expressive power of fuzzy community detection compared with hard ones. Whereas credal partition [2], which is even more general and allows in some cases to gain deeper insight into the structure of the data, it has not been applied to community detection.

In this paper, an algorithm for detecting overlapping community structure is proposed based on credal partition. An evidential modular function is introduced to determine the optimal number of communities. Spectral relaxation and evidential $c$-means are conducted to obtain the basic belief assignment (bba) of each nodes in the network. The experiments on two well-studied networks show that meaningful partitions of the graph could be obtained by the proposed detection approach and it indeed could provide us more informative information of the graph structure than the existing methods.

## 2 Background

### 2.1 Modularity-Based Community Detection

Let $G(V, E, W)$ be an undirected network, $V$ is the set of $n$ nodes, $E$ is the set of $m$ edges, and $W$ is a $n \times n$ edge weight matrix with elements $w_{ij}, i, j = 1, 2, \cdots, n$. The objective of the hard (crisp) community detection is to divide graph $G$ into $c$ clusters, denoted by

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_c\}, \tag{1}$$

and each node should belong to one and only one of the detected communities [8]. Parameter $c$ can be given in advanced or determined by the detection method itself.

The modularity, which measures the quality of a partition of a graph, was first introduced by Newman and Girvan [10]. This validity index measures how good a specific community structure is by calculating the difference between the actual edge density intra-clusters in the obtained partition and the expected one under some null models, such as random graph. One of the most popular form of modularity is given by [3]. Given a partition with $c$ group shown in Eq. (1), and let $\|W\| = \sum_{i,j=1}^{n} w_{ij}$, $k_i = \sum_{j=1}^{n} w_{ij}$, its modularity can be defined as:

$$Q_h = \frac{1}{\|W\|} \sum_{k=1}^{c} \sum_{i,j=1}^{n} (w_{ij} - \frac{k_i k_j}{\|W\|}) \delta_{ik} \delta_{jk}, \tag{2}$$

where $\delta_{ik}$ is one if vertex $i$ belongs to the $k_{th}$ community, 0 otherwise.

The communities of graph $G$ can be detected by modularity optimization, like spectral clustering algorithm [13], which aims at finding the optimal partition with the maximum modularity value [3].

### 2.2 Belief Function Theory and Evidential *c*-Means

The credal partition, a general extension of the crisp and fuzzy ones in the theoretical framework of belief function theory, has been introduced in [2,7]. Suppose the discernment frame of the clusters is $\Omega$ as in Eq. (1). Partial knowledge regarding the actual cluster node $n_i$ belongs to can be represented by a basis belief assignment defined as a function $m$ from the power set of $\Omega$ to $[0, 1]$, verifying $\sum_{A\subseteq\Omega} m(A) = 1$. Every $A \in 2^{\Omega}$ such that $m(A) > 0$ is called a focal element. The credibility and plausibility functions are defined in Eq. (3) and Eq. (4).

$$Bel(A) = \sum_{\emptyset\neq B\subseteq A} m(B), \forall A \subseteq \Omega, \tag{3}$$

$$Pl(A) = \sum_{B\cap A\neq\emptyset} m(B), \forall A \subseteq \Omega. \tag{4}$$

Each quantity $Bel(A)$ represents the degree to which the evidence supports $A$, while $Pl(A)$ can be interpreted as an upper bound on the degree of support that could be assigned to $A$ if more specific information is available [12]. The function $pl : \Omega \to [0, 1]$ such that $pl(\omega) = Pl(\{\omega\})$ is called the contour function associated to $m$.

The bbas in the credal level can be expressed in the form of probabilities by pignistic transformation [2], which is defined as

$$BetP(\omega_i) = \sum_{\omega_i\in A\subseteq\Omega} \frac{m(A)}{|A|(1 - m(\emptyset))}, \tag{5}$$

where $|A|$ is the number of elements of $\Omega$ in $A$.

Evidential $c$-means (ECM) [7] is a direct generalization of FCM. The optimal credal partition is obtained by minimizing the following objective function:

$$J_{\text{ECM}} = \sum_{i=1}^{n} \sum_{A_j\subseteq\Omega, A_j\neq\emptyset} |A_j|^{\alpha} m_i(A_j)^{\beta} d_{ij}^2 + \sum_{i=1}^{n} \delta^2 m_i(\emptyset)^{\beta}, \tag{6}$$

constrained on

$$\sum_{A_j\subseteq\Omega, A_j\neq\emptyset} m_i(A_j) + m_i(\emptyset) = 1, \tag{7}$$

where $m_i(A_j)$ is the bba of $n_i$ given to the nonempty set $A_j$, while $m_i(\emptyset)$ is the bba of $n_i$ assigned to the emptyset. The value $d_{ij}$ denotes the distance between $n_i$ and the barycenter associated to $A_j$, and $|\cdot|$ is the cardinal of the set. Parameters $\alpha, \beta, \delta$ are adjustable and can be determined based on the requirement.

## 3 Evidential Community Detection

Before presenting the credal partition of a graph $G(V, E, W)$, the hard and fuzzy partitions are firstly recalled. The crisp partition can be represented by a matrix $U^h = (u_{ik})_{n\times c}$, where $u_{ik}^h = 1$ if the $i_{th}$ node $n_i$ belongs to the $k_{th}$ cluster $\omega_i$ in the partition, and $u_{ik}^h = 0$ otherwise. From the property of this partition, it clearly should satisfy that $\sum_{k=1}^{c} u_{ik}^h = 1, i = 1, 2, \cdots, n$. The generalization of the hard partition, following that a node may belong to more communities than one but with different degrees, can be described by the fuzzy partition matrix $U^f = (u_{ik})_{n\times c}$, where $u_{ik}^f$ is not restricted in $\{0, 1\}$ but can attain any real value from the interval $[0, 1]$. The value $u_{ik}^f$ could be interpreted as a degree of membership of $n_i$ to community $\omega_k$.

The credal partition of $G$, which refers to the framework of belief function theory, can be represented by a $n$-tuple: $M = (\boldsymbol{m}_1, \boldsymbol{m}_2, \cdots, \boldsymbol{m}_n)$. Each $\boldsymbol{m}_i = \{m_{i1}, m_{i2}, \cdots, m_{i2^c}\}$ is a bba in a $2^c$-dimensional space, where $c$ is the cardinality of the given discernment frame of communities $\Omega = \{\omega_1, \omega_2, \cdots, \omega_c\}$ as before, and $\omega_i$ denotes the $i_{th}$ detected community. Note that $\Omega$ is the discernment frame in the framework of belief function theory.

### 3.1 The Evidential Modular Function

Similar to the fuzzy modularity by Nepusz et al. [8] and by Havens et al. [5], here we introduce an evidential modularity:

$$Q_e = \frac{1}{\|W\|} \sum_{k=1}^{c} \sum_{i,j=1}^{n} (w_{ij} - \frac{k_i k_j}{\|W\|}) pl_{ik} pl_{jk}, \tag{8}$$

where $\boldsymbol{pl}_i = \{pl_{i1}, pl_{i2}, \cdots, pl_{ic}\}$ is the contour function associated to $m_i$, which describes the upper value of our belief to the proposition that the $i_{th}$ node belongs to the $k_{th}$ community.

Let $\boldsymbol{k} = (k_1, k_2, \cdots, k_n)^T$, $B = W - \boldsymbol{k}^T\boldsymbol{k}/\|W\|$, and $PL = (pl_{ik})_{n\times c}$, then Eq. (8) can be rewritten as:

$$Q_e = \frac{\text{trace}(PL\ B\ PL^T)}{\|W\|}. \tag{9}$$

$Q_e$ is a directly extension of the crisp modularity function (2). When the credal partition degrades into the hard one, $Q_e$ is equal to $Q_h$.

### 3.2 Spectral Mapping

White and Smyth [13] showed that optimizing the modularity measure $Q$ can be reformulated as a spectral relaxation problem and proposed spectral clustering algorithms that seek to maximize $Q$. By eigendecomposing a related matrix, these methods can map graph data points into Euclidean space, the clustering problem on which space is of equivalence to that on the original graph.

Let $A = (a_{ij})_{n\times n}$ be the adjacent matrix of the graph $G$. The adjacency matrix for a weighted graph is given by the matrix whose element $a_{ij}$ represents the weight $w_{ij}$ connecting nodes $i$ and $j$. The degree matrix $D = (d_{ii})$ is the diagonal matrix whose elements are the degrees of the nodes of $G$, *i.e.* $d_{ii} = \sum_{j=1}^{n} a_{ij}$. The eigenvectors of the transition matrix $\mathcal{M} = D^{-1}A$ are used.

Verma and Meila [14] and Zhang et al. [17] suggested to use the eigenvectors of a generalised eigensystem $Ax = \lambda Dx$, and pointed out that it is mathematically equivalent and numerically more stable than computing the eigenvectors of matrix $\mathcal{M}$ [14]. To partition the nodes of the graph into $c$ communities, the top $c-1$ eigenvectors of the above eigensystem are used to map the graph data into points in the Euclidean space, where the traditional clustering methods, such as $c$-means (CM), FCM and ECM can be evoked.

### 3.3 Evidential Community Detection Scheme

Let $C$ be the upper bound of the number of communities. The evidential community detection scheme is displayed as follows:

**S.1** Spectral mapping:
For $2 \leq c \leq C$, Find the top $c$ generalized eigenvectors $E_c = [e_1, e_2, \cdots, e_c]$ of the eigensystem $Ax = \lambda Dx$, where $A$ and $D$ are the adjacent and the degree matrix respectively.

**S.2** Evidential $c$-means:
For each value of $c$ ($2 \leq c \leq C$), let $E_c = [e_2, \cdots, e_c]$. Use ECM to partition the $n$ samples (each row of $E_c$ is a sample data on the $c-1$ dimensional Euclidean space) into $c$ classes. And we can get a credal partition $M$ for the graph.

**S.3** Choosing the number of communities:
Find the suitable number of clusters and the corresponding evidential partition scheme by maximizing the evidential modular function $Q_e$.

In the algorithm, $C$ can be determined by the original graph. It is an empirical range of the community number of the network. If $c$ is given, we can get a credal partition using the proposed method and then the evidential modularity can be derived. The modularity is a function of $c$ and it should peak around the optimum value of $c$ for the given network. As in ECM, the number of parameters to be optimized is exponential in the number of communities and linear in the number of nodes. When the number of communities is large, we can reduce the complexity by considering only a subclass of bbas with a limited number of focal sets [7].

## 4 Experimental Results

To evaluate the proposed method in this paper, two real-world networks are discussed in this section. A comparison for the detected communities by credal, hard and fuzzy partitions is also illustrated to show the advantages of evidential community structure over others.

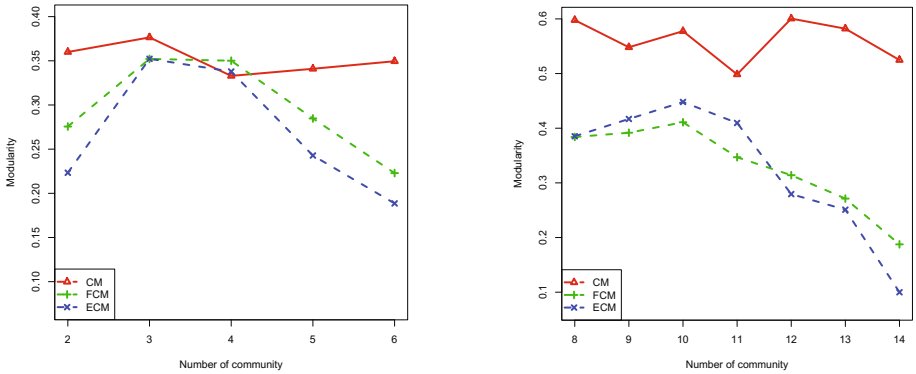
### 4.1 Zachary's Karate Club

The Zachary's Karate Club [16] is an undirected graph which consists of 34 vertices and 78 edges, describing the friendship between members of the club observed by Zachary in his two-year study. This club is visually divided into two parts, due to an incipient conflict between the president and instructor (see Fig. 2-a).

The modularity peaks around $c = 2$ or $c = 3$ as shown in Fig. 1-a. Let $c = 3$, the detected communities by CM, FCM and ECM are displayed in Fig. 2. As it can be seen, a small community separated from $\omega_1$ is detected by all the approaches. The result by FCM shown here is got by partitioning nodes to the cluster with the highest membership. Zhang et al. [17] suggested to use a threshold $\lambda$ to covert the fuzzy membership into the final community structure. For node $i$, let the fuzzy assignment to its communities be $\mu_{ij}, j = 1, 2, \cdots, c$. Node $i$ is regarded as a member of multiple communities $\omega_k$ with $\mu_{ik} > \lambda$. But there is no criterion for determining the appropriate $\lambda$. However, in ECM we can directly get the imprecise classes indicating our uncertainty on the actual cluster of some nodes by hard credal partitions [7].

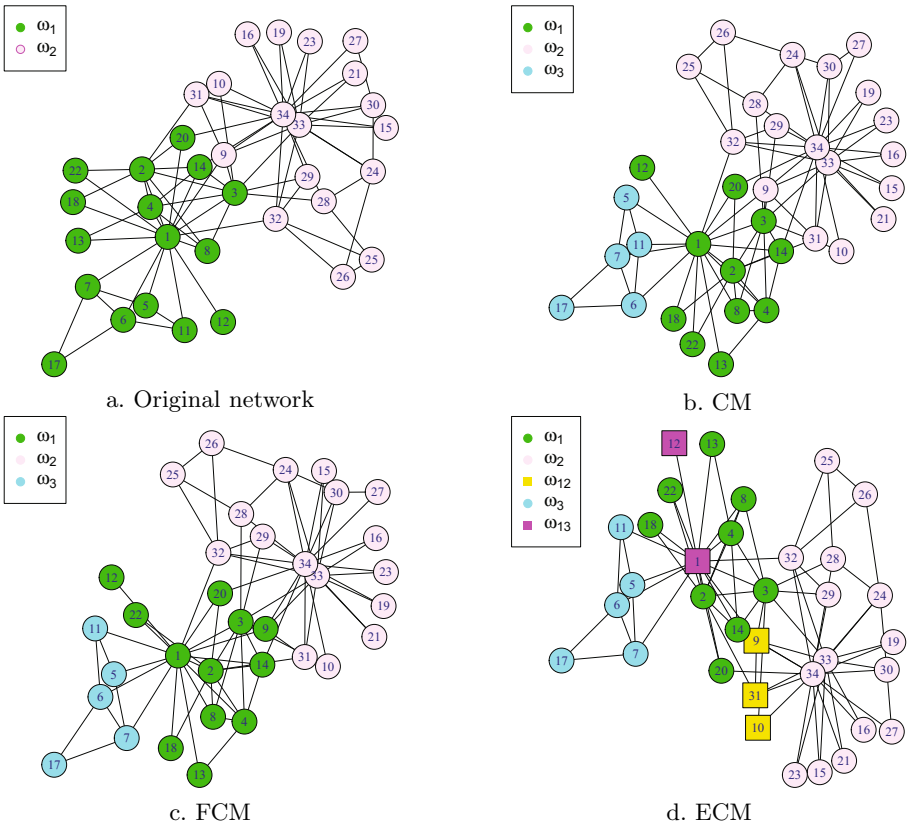
As we can see in Fig. 2-c, for ECM, node 1,9,10,12,31 belong to two clusters at the same time. This is coincident with the conclusion in [17] apart from the fact that a significant high membership value is given to $\omega_1$ for node 12 by FCM. Actually, the case that node 12 is clustered into $\omega_{12} \triangleq \{\omega_1, \omega_2\}$ seems reasonable when the special behavior of this node is considered. The person 12 has no contact with others except the instructor (node 1). Therefore, the most probable class of node 12 should be the same as that of node 1. It is counterintuitive if the person 12 is partitioned into either $\omega_1$ or $\omega_2$, as it has no relation with any member in these two communities at all. The credal partition can reflect the fact that $\omega_1$ and $\omega_2$ is indistinguishable to node 12, while the fuzzy method could not. Furthermore, the mass belief assigned to imprecise classes reflects our degree of uncertainty on the clusters of the included nodes. As illustrated in Fig. 3-b, the mass given to imprecise clusters for node 1 is larger than that to the other four nodes. This reflects our uncertain on node 1's community is largest. As node 1 is the instructor of the club, this fact seems reasonable.

Actually, the concept of credal partitions suggests different ways of summarizing data. For example, the data can be analysed in the form of fuzzy partition thanks to the pignistic probability transformation shown in Eq. (5). It is shown in Fig. 3-a pignistic probabilities play the same role as fuzzy membership. A crisp partition can then be easily obtained by partitioning each node to the community with the highest pignistic probability. In this sense, the proposed method could be regarded as a general model of hard and fuzzy community detection approaches.




a. Karate Club network b. American football network

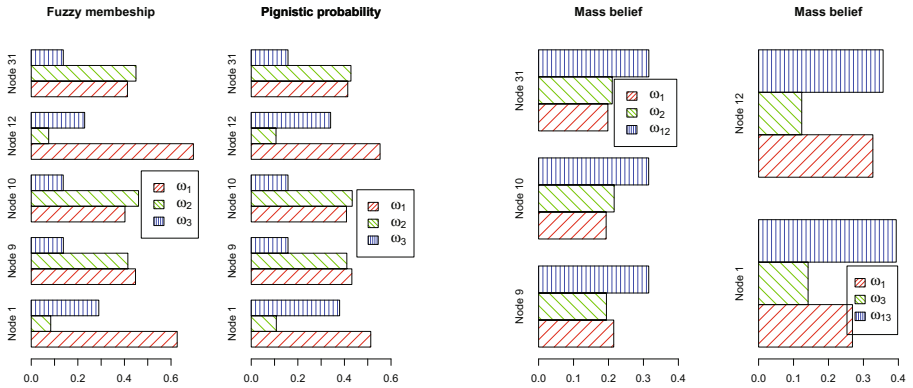
**Fig. 1.** Modularity values with community numbers




a. Original network b. CM

c. FCM d. ECM

**Fig. 2.** Karate Club




a. Probability membership b. Mass belief

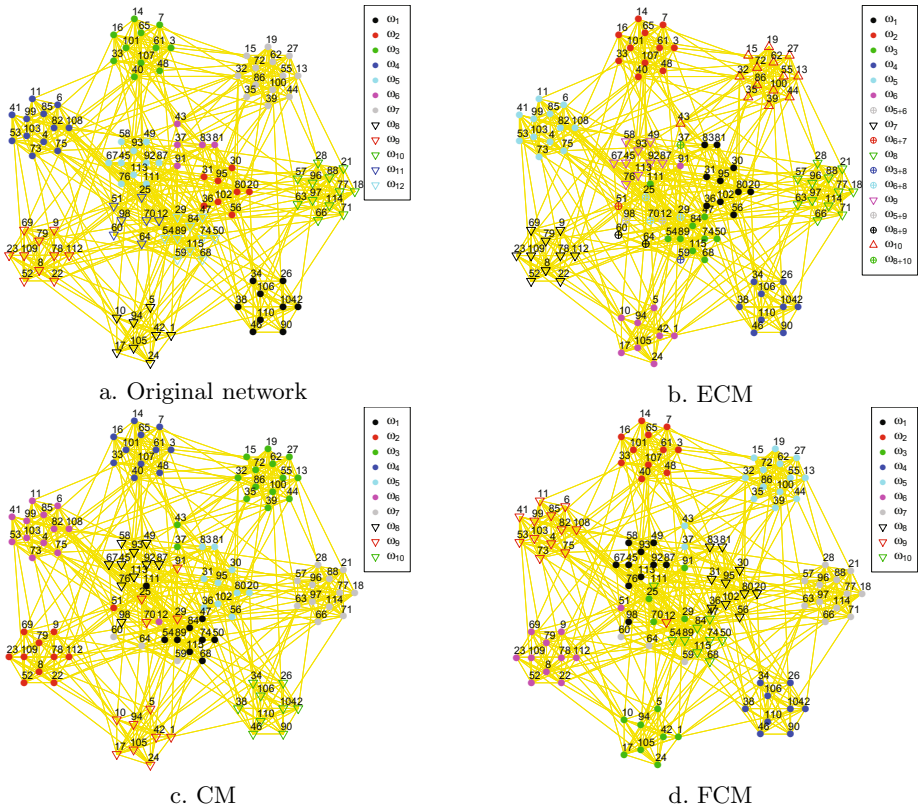
**Fig. 3.** Clustering results of Karate club network

### 4.2 American Football Network

The network we investigate in this experiment is the world of American college football games between Division IA colleges during regular season Fall 2000 [4]. The vertices in the network represent 115 teams, while the links denote 613 regular-season games between the two teams they connect. The teams are divided into 12 conferences containing around 8-12 teams each and generally games are more frequent between members from the same conference than between those from different conferences.

In ECM, the number of parameters to be optimized is exponential in the number of clusters [7]. For the number of class larger than 10, calculations are not tractable. But we can consider only a subclass with a limited number of focal sets [7]. In this example, we constrain the focal sets to be composed of at most two classes (except $\Omega$). Fig. 1-b shows how the modularity varies with the number of communities. For credal partitions, the peak is at $c = 10$. This is consensus with the original network (shown in Fig. 4-a) composed of 10 large communities (more than 8 members) and 2 small communities (8 members or less than 8 members). Set $c = 10$ in ECM, we can find all the ten large communities, eight of which are exactly detected. For the nodes in small communities, ECM partitions most of them into imprecise classes. As there are more than 10 communities in this network, we use $\omega_{i+j}$ to denote the imprecise communities instead of $\omega_{ij}$ in the figures related to this experiment to obviate misunderstanding.

For hard partitions, nodes in small communities are simply partitioned into their "closest" detected cluster, which will certainly result in a loss of accuracy for the final results. Credal partitions make cautious decisions by clustering nodes which we are uncertain into imprecise communities. The introduced imprecise clusters can avoid the risk to group a node into a specific class without strong belief. In other words, a data pair can be clustered into the same specific group only when we are quite confident and thus the misclassification rate will be reduced.



a. Original network b. ECM c. CM d. FCM

**Fig. 4.** American football network

## 5 Conclusion

In this paper, a new community detection approach combing the evidential modularity, spectral mapping and evidential $c$-means is presented to identify the overlapping graph structure in complex networks. Although many overlapping community-detection algorithms have been developed before, most of them are based on fuzzy partitions. Credal partitions, in the frame of belief function theory, have many advantages compared with fuzzy ones and enable us to have a better insight into the data structure. As shown in the experimental results for two networks in the real world, credal partitions can reflect our degree of uncertain more intuitively. Actually, the credal partition is an extension of both hard and fuzzy ones, thus there is no doubt that more rich information of the graph structure could be available from the detected structure by the method proposed here. We expect that the evidential clustering approaches will be employed with promising results in the detection of overlapping communities in complex networks with practical significance.

## References


1. Costa, L.D.F., Oliveira Jr., O.N., Travieso, G., Rodrigues, F.A., Villas Boas, P.R., Antiqueira, L., Viana, M.P., Correa Rocha, L.E.: Analyzing and modeling real-world phenomena with complex networks: a survey of applications. Advances in Physics 60(3), 329–412 (2011)
2. Denœux, T., Masson, M.H.: Evclus: evidential clustering of proximity data. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34(1), 95–109 (2004)
3. Fortunato, S.: Community detection in graphs. Physics Reports 486(3), 75–174 (2010)
4. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99(12), 7821–7826 (2002)
5. Havens, T., Bezdek, J., Leckie, C., Ramamohanarao, K., Palaniswami, M.: A soft modularity function for detecting fuzzy communities in social networks. IEEE Transactions on Fuzzy Systems 21(6), 1170–1175 (2013)
6. Havens, T.C., Bezdek, J.C., Leckie, C., Chan, J., Liu, W., Bailey, J., Ramamohanarao, K., Palaniswami, M.: Clustering and visualization of fuzzy communities in social networks. In: 2013 IEEE International Conference on Fuzzy Systems (FUZZ), pp. 1–7. IEEE (2013)
7. Masson, M.H., Denoeux, T.: Ecm: An evidential version of the fuzzy c-means algorithm. Pattern Recognition 41(4), 1384–1397 (2008)
8. Nepusz, T., Petróczi, A., Négyessy, L., Bazsó, F.: Fuzzy communities and the concept of bridgeness in complex networks. Physical Review E 77(1), 016107 (2008)
9. Newman, M.E.: Fast algorithm for detecting community structure in networks. Physical review E 69(6), 066133 (2004)
10. Newman, M.E., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E 69(2), 026113 (2004)
11. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043), 814–818 (2005)
12. Smets, P., Kennes, R.: The transferable belief model. Artificial Intelligence 66(2), 191–234 (1994)
13. Smyth, S., White, S.: A spectral clustering approach to finding communities in graphs. In: Proceedings of the 5th SIAM International Conference on Data Mining, pp. 76–84 (2005)
14. Verma, D., Meila, M.: A comparison of spectral clustering algorithms. Tech. rep., UW CSE (2003)
15. Wang, X., Jiao, L., Wu, J.: Adjusting from disjoint to overlapping community detection of complex networks. Physica A: Statistical Mechanics and its Applications 388(24), 5045–5056 (2009)
16. Zachary, W.W.: An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 452–473 (1977)
17. Zhang, S., Wang, R.S., Zhang, X.S.: Identification of overlapping community structure in complex networks using fuzzy c-means clustering. Physica A: Statistical Mechanics and its Applications 374(1), 483–490 (2007)

# Probabilistic Fuzzy Systems as Additive Fuzzy Systems

Rui Jorge Almeida[1], Nick Verbeek[2], Uzay Kaymak[1], and João Miguel Da Costa Sousa[3]

[1] School of Industrial Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
rjalmeida@tue.nl, u.kaymak@ieee.org
[2] Erasmus University Rotterdam, Erasmus School of Economics, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
291847nv@student.eur.nl
[3] Universidade de Lisboa, Instituto Superior Tecnico, Dept. of Mechanical Engineering, CSI, IDMEC/LAETA Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
jmsousa@tecnico.ulisboa.pt

**Abstract.** Probabilistic fuzzy systems combine a linguistic description of the system behaviour with statistical properties of data. It was originally derived based on Zadeh's concept of probability of a fuzzy event. Two possible and equivalent additive reasoning schemes were proposed, that lead to the estimation of the output's conditional probability density. In this work we take a complementary approach and derive a probabilistic fuzzy system from an additive fuzzy system. We show that some fuzzy systems with universal approximation capabilities can compute the same expected output value as probabilistic fuzzy systems and discuss some similarities and differences between them. A practical relevance of this functional equivalence result is that learning algorithms, optimization techniques and design issues can, under certain circumstances, be transferred across different paradigms.

**Keywords:** Probabilistic Fuzzy Systems, Additive Fuzzy Systems, Conditional Density Approximation.

## 1 Introduction

Probabilistic fuzzy systems (PFS) can deal explicitly and simultaneously with fuzziness or linguistic uncertainty and probabilistic uncertainty. A probabilistic fuzzy system follows an idea similar to [1–4] where the different concepts [5–8] of fuzzy sets and probabilities are complementary [6].

As a mathematical notion, a fuzzy set $F$ on a finite universe $U$ is unambiguously defined by a membership function $u_F : U \rightarrow [0, 1]$. The mathematical object representing the fuzzy set is the membership function $u_F(x)$ indicating the grade of membership of element of $x \in U$ in $F$. At the mathematical level the domain of the fuzzy sets is $[0, 1]^U$. On the other hand a probabilistic measure Pr

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 567–576, 2014.

of an experiment $\epsilon$ yet to be performed, is a mapping $2^U \rightarrow [0,1]$ that assigns a number $\Pr(A)$ of event $A$ to each subset of $U$, satisfying the Kolmogorov axioms. $\Pr(A)$ is the probability that a generic outcome of $\epsilon$, an ill-known single-valued variable $x$, hits the well-known set $A$. If the outcome of $\epsilon$ is such that $x \in A$, then we say that event $A$ has occurred. In this case there is uncertainty about the occurrence of any particular $x$ and consequently of event $A$. This uncertainty is described by $\Pr(A)$. At the mathematical level the domain of the mapping Pr is the Boolean algebra $2^U$.

Various rule base structures and reasoning mechanisms for fuzzy systems (e.g. [9–11]), emphasize the modelling of the linguistic uncertainty and interpolation capability of fuzzy systems, being typically used for approximating deterministic functions, in which the stochastic uncertainty is ignored. A probabilistic fuzzy system, as it was formally defined in [12], was based on the concept of probability of fuzzy events. This type of system estimates a conditional probability density function for the output variable, given the inputs to the system. Two equivalent additive reasoning mechanism have been proposed for PFS, one based on the concept of fuzzy histograms and another based on the stochastic mapping between fuzzy antecedents and fuzzy consequents.

In this work we follow a different reasoning and derive a probabilistic fuzzy system starting from a additive fuzzy system. This different analysis provides a different insight and understanding of probabilistic fuzzy systems, which can be related to Mamdani fuzzy systems and fuzzy relational models and departs from the concept of probability of fuzzy events. This allows us to formalize the definition of probabilistic fuzzy systems while exposing similarities and differences with different models or concepts. The relation of probabilistic fuzzy system to well known fuzzy systems helps to explain its success for function approximation. A practical relevance of the functional equivalence presented in this work is that learning algorithms, optimization techniques and design issues can be transferred to probabilistic fuzzy systems. Furthermore, it also allows to interpret models transversely across different modeling paradigms.

The outline of the paper is as follows. In Section 2, we give an overview of the original definition of probabilistic fuzzy systems and present the two equivalent additive reasoning mechanisms of a PFS, as well as the different outputs. In Section 3 we present the new derivation of a probabilistic fuzzy system starting from fuzzy additive systems and discuss in Section 4 several issues related to our findings. Finally we conclude the paper in Section 5.

## 2 Probabilistic Fuzzy Systems

Probabilistic fuzzy systems combine two different types of uncertainty, namely fuzziness or linguistic vagueness, and probabilistic uncertainty. In this work we consider that the probabilistic uncertainty relate to aleatoric variability, while fuzzy sets are used to represent gradualness, epistemic uncertainty or bipolarity [7, 13].

The PFS consists of a set of rules whose antecedents are fuzzy conditions and whose consequents are probability distributions. Assuming that the input

space is a subset of $\mathbb{R}^n$ and that the rule consequents are defined on a finite domain $Y \subseteq \mathbb{R}$, a probabilistic fuzzy system consists of a system of rules $R_q$, $q = 1, \ldots, Q$, of the type

$$R_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f(y) \text{ is } f(y|A_q)\,, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is an input vector, $A_q : X \longrightarrow [0,1]$ is a fuzzy set defined on $X$ and $f(y|A_q)$ is the conditional pdf of the stochastic output variable $\underline{y}$ given the fuzzy event $A_q$. The interpretation is as follows: if fuzzy antecedent $A_q$ is fully valid ($x \in \text{core}(A_q)$), then $y$ is a sample value from the probability distribution with conditional pdf $f(y|A_q)$.

A PFS has been described with two possible and equivalent reasoning mechanisms, namely the fuzzy histogram approach and the probabilistic fuzzy output approach [12]. In both cases, we suppose that $J$ fuzzy classes $C_j$ form a fuzzy partition of the compact output space $Y$.

### 2.1 Fuzzy Histogram Model

In the fuzzy histogram approach, we replace in each rule of (1) the true pdf $f(y|A_q)$ by its fuzzy approximation (fuzzy histogram) $\hat{f}(y|A_q)$ yielding the rule set $\hat{R}_q$, $q = 1, \ldots, Q$ defined as

$$\hat{R}_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f(y) \text{ is } \hat{f}(y|A_q)\,, \tag{2}$$

where $\hat{f}(y|A_q)$ is a fuzzy histogram conform [14]

$$\hat{f}(y|A_q) = \sum_{j=1}^{J} \frac{\hat{\Pr}(C_j|A_q) u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}\,. \tag{3}$$

The numerator in (3) describes a superposition of fuzzy events described by their membership functions $u_{C_j}(y)$, weighted by the probability $\hat{\Pr}(C_j|A_q)$ of the fuzzy event. The denominator of (3) is a normalizing factor representing the fuzzified size of class $C_j$. Because of overlapping membership functions, fuzzy histograms have a high level of statistical efficiency, compared to crisp ones and several classes of fuzzy histograms also have a high level of computational efficiency [15].

The interpretation of this type of reasoning is as follows. Given the occurrence of a (multidimensional) antecedent fuzzy event $A_q$, which is a conjunction of the fuzzy conditions defined on input variables, an estimate of the conditional probability density function based on a fuzzy histogram $\hat{f}(y|A_q)$ is calculated.

### 2.2 Probabilistic Fuzzy Output Model

In the probabilistic fuzzy output approach, sometimes also referred to as Mamdani PFS [16–18], we decompose each rule (1) to provide a stochastic mapping

between its fuzzy antecedents and its fuzzy consequents. The rules are written in the following form.

$$\begin{aligned} \text{Rule } \hat{R}_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then } \quad & \underline{y} \text{ is } C_1 \text{ with } \hat{\Pr}(C_1|A_q) \text{ and} \\ & \cdots \\ & \underline{y} \text{ is } C_J \text{ with } \hat{\Pr}(C_J|A_q). \end{aligned} \tag{4}$$

These rules specify a probability distribution over a collection of fuzzy sets that partition the output domain. The rules of a PFS also express linguistic information and they can be used to explain the model behaviour by a set of linguistic rules. This way, the system deals both with linguistic uncertainty as well as probabilistic uncertainty.

The interpretation for the probabilistic fuzzy output approach is as follows. Given the occurrence of a (multidimensional) antecedent fuzzy event $A_q$, which is a conjunction of the fuzzy conditions defined on input variables, each of the consequent fuzzy events $C_j$ is likely to occur. The selection of consequent fuzzy events is done proportionally to the conditional probabilities $\hat{\Pr}(C_j|A_q)$, $(j = 1, 2, \ldots, J)$. This applies for all the rules $R_q$, $q = 1, 2, \ldots, Q$.

The probabilistic fuzzy system in this form resembles a deterministic Mamdani fuzzy system with rule base multiplicative implication and additive aggregation. The difference lies in the fact that in a Mamdani fuzzy system only one of the outputs is considered in each rule, while in a PFS, each fuzzy output $C_q$ can happen with a given conditional probability $\hat{\Pr}(A_q|C_j)$.

### 2.3 Outputs of Probabilistic Fuzzy Systems

Although the fuzzy rule bases (2) and (4) are different, under certain conditions, the two corresponding probabilistic fuzzy systems implement the same crisp input-output mapping [12]. The output of the fuzzy rules (4) is the same as in the rules (2), if an additive reasoning scheme is used with multiplicative aggregation of the rule antecedents [19].

Given an input vector $\mathbf{x}$, the output of a probabilistic fuzzy system is a conditional density function which can be computed as

$$\hat{f}(y|\mathbf{x}) = \sum_{j=1}^{J} \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \hat{\Pr}(C_j|A_q) \frac{u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}, \tag{5}$$

where

$$\beta_q(\mathbf{x}) = \frac{u_{A_q}(\mathbf{x})}{\sum_{q'=1}^{Q} u_{A_{q'}}(\mathbf{x})} \tag{6}$$

is the normalised degree of fulfillment of rule $R_q$ and $u_{A_q}$ is the degree of fulfillment of rule $R_q$. When $\mathbf{x}$ is $n$-dimensional, $u_{A_q}$ is determined as a conjunction of the individual memberships in the antecedents computed by a suitable t-norm, *i.e.*, $u_{A_q}(\mathbf{x}) = u_{A_{q_1}}(x_1) \circ \cdots \circ u_{A_{q_n}}(x_n)$, where $x_i, i = 1, \ldots, n$ is the $i$-th

component of $\mathbf{x}$ and $\circ$ denotes a t-norm. From the obtained output probability distribution it is possible to calculate a crisp output using the expected value

$$\hat{\mu}_{y|\mathbf{x}} = \hat{E}(y|\mathbf{x}) = \int_{-\infty}^{\infty} y\hat{f}(y|\mathbf{x})dy = \sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\Pr}(C_j|A_q)z_{1,j}\,, \tag{7}$$

where $z_{1,j}$ is the centroid of the $j$th output fuzzy set defined by

$$z_{1,j} = \frac{\int_{-\infty}^{\infty} y u_{C_j}(y)dy}{\int_{-\infty}^{\infty} u_{C_j}(y)dy}\,. \tag{8}$$

It can be shown [12] that the conditional density output $\hat{f}(y|\mathbf{x})$ of a PFS is a proper probability density function *i.e.,* $\int_{-\infty}^{\infty} \hat{f}(y|\mathbf{x})dy = 1$ and that the crisp outputs, expected value $\hat{\mathrm{E}}(\underline{y}|\mathbf{x})$ and second moment $\hat{\mathrm{E}}(\underline{y}^2|\mathbf{x})$, exist if the output space is well-formed, *i.e.* the output membership values satisfy

$$\sum_{j=1}^{J} u_{C_j}(y) = 1, \qquad \forall y \in Y, y < \infty. \tag{9}$$

## 3 Probabilistic Fuzzy Systems as Additive Fuzzy Systems

In this work we depart from the previous definition presented in Section 2.2 and instead derive a probabilistic fuzzy system from an additive fuzzy system. This deterministic fuzzy system has rule base multiplicative implication and additive aggregation, where the crisp model output is obtained using the center of gravity defuzzification method. In the following section we present the additive fuzzy system under consideration and how it can be converted in a probabilistic fuzzy system.

### 3.1 Additive Fuzzy Systems

Let $R = \cup_{q=1}^{Q} R_q$ be a rule base for a additive fuzzy system of the type

$$\begin{aligned} \text{Rule } \hat{R}_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then } & \underline{y} \text{ is } C_1 \text{ with } w(A_q, C_1) \text{ and} \\ & \dots \\ & \underline{y} \text{ is } C_J \text{ with } w(A_q, C_J)\,, \end{aligned} \tag{10}$$

where $w(A_q, C_j) \in \mathbb{R}_{\geq 0}$ are non-negative weights. The system defined by (10) is similar to the standard additive model [20, 21] but in the former, the consequents are not directly dependent on $x$.

Although the fuzzy rule base system defined by (10) can be obtained by replacing the conditional probabilities $\hat{\Pr}(C_j|A_q)$ by non-negative weights $w(A_q, C_j) \in \mathbb{R}_{\geq 0}$ in the fuzzy rule system (4), the crisp output of both systems is different, as the following theorem shows.

**Theorem 1.** *Let $R = \cup_{q=1}^{Q} R_q$ be a fuzzy rule base as defined by (10) such that $u_{A_q}(\mathbf{x}) > 0, \forall q$ and the output space follows (9), and the rule base uses multiplicative implication and additive aggregation. Then the crisp model output $y^*$ obtained using the center of gravity defuzzification method is*

$$y^* = \frac{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) v_{1,j} z_{1,j}}{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) v_{1,j}}, \tag{11}$$

*where $z_{1,j}$ is given by (8) and $v_{1,j}$ is the area of the $j$th output fuzzy set defined by*

$$v_{1,j} = \int_{-\infty}^{\infty} u_{C_j}(y) dy. \tag{12}$$

*Proof.* The center of gravity defuzzification method is given by

$$y^* = \frac{\int_{-\infty}^{\infty} y \chi(x,y) dy}{\int_{-\infty}^{\infty} \chi(x,y) dy}, \tag{13}$$

where $\chi(x,y)$ is the output of the fuzzy system under consideration. For the case of the additive fuzzy system (10) using with multiplicative implication and additive aggregation the output is

$$\chi(x,y) = \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) u_{C_j}(y). \tag{14}$$

Substituting (14) into (13) we obtain

$$\begin{aligned} y^* &= \frac{\int_{-\infty}^{\infty} y \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q,C_j) u_{C_j}(y) dy}{\int_{-\infty}^{\infty} \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q,C_j) u_{C_j}(y) dy} \\ &= \frac{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q,C_j) \int_{-\infty}^{\infty} y u_{C_j}(y) dy}{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q,C_j) \int_{-\infty}^{\infty} u_{C_j}(y) dy} \\ &= \frac{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q,C_j) \int_{-\infty}^{\infty} u_{C_j}(y) dy \frac{\int_{-\infty}^{\infty} y u_{C_j}(y) dy}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}}{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q,C_j) \int_{-\infty}^{\infty} u_{C_j}(y) dy} \\ &= \frac{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q,C_j) v_{1,j} z_{1,j}}{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q,C_j) v_{1,j}}. \end{aligned} \tag{15}$$

### 3.2 Equivalence to Probabilistic Fuzzy Systems

Starting from an additive fuzzy system defined in (10), it is possible to obtain a probabilistic fuzzy system. Before formalizing this result we introduce the following definition of a probability kernel.

**Definition 1.** *A kernel is a mapping $\mathcal{K} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$ from the measurable space $(\mathbf{X}, \mathcal{X})$ to the measurable space $(\mathbf{Y}, \mathcal{Y})$. The kernel $\mathcal{K}$ is a probability kernel if it is defined as a probability measure on $(\mathbf{Y}, \mathcal{Y})$.*

Given this definition we can now prove that a probabilistic fuzzy system can be obtained starting from the fuzzy system defined in (10).

**Theorem 2.** *If the mapping $w(A_q, C_j)$ is defined as a probability kernel and each output consequent $C_j$ are functions defined on a random variable space then the output of the PFS is a conditional probability density for $y$ given $x$. Under this definition, the fuzzy rule base in (10) has a functional equivalent to the PFS in (4) and the crisp output (11) has a functional equivalent to the conditional output of the PFS in (5).*

*Proof.* The defined non-negative weights $w(A_q, C_j) : (\mathbf{X} \times Y) \rightarrow R_{\geq 0}$ form a kernel on the measurable space $(\mathbb{R}^n \times \mathbb{R})$. If $w(A_q, C_j)$ is also defined as a probability measure on $(Y, \mathcal{Y})$, such that $\sum_{j=1}^{J} w(A_q, C_j) = 1, \forall q = 1, \ldots, Q$ then according to Definition 1, $w(A_q, C_j)$ is a probability kernel. We recall that using (6) we obtain $\sum_{q=1}^{Q} \beta_q(\mathbf{x}) = 1$. Furthermore, since the output fuzzy sets $C_j$ are admissible functions for defining random variables then they are limited to those for which a probability distribution exists. A simple form to ensure this is to normalize them

$$u_{C_j'} = \frac{u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy}\,. \tag{16}$$

The output of the fuzzy system $\chi(x, y)$ in (13) is then a conditional density function for $Y$ given $X$ such that:

$$\begin{aligned} \int_{-\infty}^{\infty} \chi(x,y)dy &= \int_{-\infty}^{\infty} \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) u_{C_j'} dy \\ &= \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) \frac{\int_{-\infty}^{\infty} u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} = 1\,. \end{aligned} \tag{17}$$

In the case that $w(A_q, C_j)$ is defined as a probability kernel, the additive fuzzy system defined by the rule base (10) is a probabilistic fuzzy system as presented in (4). Furthermore, the center of gravity output (11) of the additive fuzzy system has a functional equivalent to the expectation of the conditional output of the PFS (5)

$$\begin{aligned} y^* &= \frac{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) v_{1,j} z_{1,j}}{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) v_{1,j}} \\ &= \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) z_{1,j}. \end{aligned} \tag{18}$$

Since $w(A_q, C_j)$ is a probability kernel, (18) is equivalent to (7).

A practical relevance of the functional equivalence result is that learning algorithms, optimization techniques and design issues can be transferred across different paradigms. Furthermore, this result helps to explain the success of fuzzy systems for function approximation in the presence of probabilistic uncertainty.

## 4 Discussion

The previous sections have shown that a probabilistic fuzzy system defined by (4) can be obtained starting from a additive fuzzy system (10). An important aspect is that since $w(A_q, C_j)$ is defined as a probability kernel then it has a

functional equivalent to $\Pr(C_j|A_q)$ in (4), implying that $\sum_{j=1}^{J} \hat{\Pr}(C_j|A_q) = 1$ and $\hat{\Pr}(C_j|A_q) \geq 0$. In this paper we do not assume any particular algebraic structure for the conditional probability of fuzzy events. There are several examples of definitions of conditional probabilities of fuzzy events that satisfy the classical axioms of conditional probabilities as given by [22] and [23] that can be found in [24–26]. This is an important issue that needs to be studied closely in the future.

It is also interesting to note that the system defined by (10) can be transformed in a fuzzy relational model [11] when $w(A_q, C_j)$ is replaced by the fuzzy relation $u(A_q, C_j)$. Similarly to a fuzzy relational model, a probabilistic fuzzy system can also be fine tuned by modifying the probability parameters $\hat{\Pr}(C_j|A_q)$, while maintaining the fuzzy input and fuzzy output space constant. We stress that a fuzzy relational model and a probabilistic fuzzy system have different interpretations, based on the nature of the uncertainty of the relation and output being modelled, as described in Section 1. In a fuzzy relational model the elements of the relation represent the strength of association between the fuzzy sets, while in the case of a fuzzy probabilistic model they are a stochastic mapping between fuzzy sets. Furthermore, the output fuzzy sets of a probabilistic fuzzy system are defined in the space of a stochastic variable $\underline{y}$. These differences leads to different nature of outputs, albeit under certain circumstances, there is a functional equivalence between both models crisp output. In the general case that $w(A_q, C_j)$ are non-negative weights, or in the case of a fuzzy relational model $u(A_q, C_j)$ are fuzzy relations, the output of such a system is not a proper probability density function.

As a result of theorem 1 and theorem 2, a Mamdani fuzzy model can be regarded as a special case of the fuzzy system defined in (10), or equivalently the system defined by (4). A Mamdani fuzzy model is recovered when the system is purely deterministic by setting setting for all $q = 1, \ldots, Q$, $\exists \kappa \in \{1 \ldots, J\}$ such that $\hat{\Pr}(C_\kappa|A_q) = 1$ and $\hat{\Pr}(C_j|A_q) = 0$, $j \neq \kappa$ *i.e.,* only one of the possible consequents is certain for each rule $Q$.

## 5 Conclusions

This paper presents a new form to derive a probabilistic fuzzy system starting from an additive fuzzy system. This new reasoning departs from the original derivation of a PFS which was based on Zadehs' concept of probability of a fuzzy event. We show that in certain cases an additive fuzzy system can compute the same expected output value as a PFS. We discuss some similarities between Mamdani and fuzzy relation models with probabilistic fuzzy systems. A practical relevance of the functional equivalence result is that learning algorithms, optimization techniques and design issues can be transferred across different paradigms. Furthermore, our results provide insight why additive deterministic fuzzy systems have proven to be so successful for function approximation purposes.

**Acknowledgements.** The authors would like to thank Didier Coquin for the discussions and help provided in this work. This work was supported by French National Agency for Research with the reference ANR-10-CORD-005 (ReVeS project).

## References

1. Meghdadi, A.H., Akbarzadeh-T., M.R.: Probabilistic fuzzy logic and probabilistic fuzzy systems. In: Proceedings of the Tenth IEEE International Conference on Fuzzy Systems, Melbourne, Australia, pp. 1127–1130 (2001)
2. Liu, Z., Li, H.X.: A probabilistic fuzzy logic system for modeling and control. IEEE Transactions on Fuzzy Systems 13, 848–859 (2005)
3. Zhang, G., Li, H.X.: A probabilistic fuzzy learning system for pattern classification. In: 2010 IEEE International Conference on Systems Man and Cybernetics (SMC), pp. 2336–2341 (2010)
4. Zhang, G., Li, H.X.: An efficient configuration for probabilistic fuzzy logic system. IEEE Transactions on Fuzzy Systems 20, 898–909 (2012)
5. Zadeh, L.A.: Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications 23, 421–427 (1968)
6. Zadeh, L.A.: Discussion: Probability theory and fuzzy logic are complementary rather than competitive. Technometrics 37, 271–276 (1995)
7. Dubois, D., Prade, H.: The three semantics of fuzzy sets. Fuzzy Sets and Systems 90, 141–150 (1997)
8. Bertoluzza, C., Gil, M.A., Ralescu, D.A. (eds.): Statistical Modeling, Analysis and Management of Fuzzy Data. STUDFUZZ. Physica Verlag, Heidelberg (2002)
9. Mamdani, E.H., Gaines, B.R. (eds.): Fuzzy Reasoning and its Applications. Academic Press, London (1981)
10. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modelling and control. IEEE Transactions on Systems, Man and Cybernetics 15, 116–132 (1985)
11. Pedrycz, W.: Applications of fuzzy relational equations for methods of reasoning in presence of fuzzy data. Fuzzy Sets and Systems 16, 163–175 (1985)
12. van den Berg, J., Kaymak, U., Almeida, R.J.: Conditional density estimation using probabilistic fuzzy systems. Accepted for Publication in IEEE Transactions on Fuzzy Systems (2013)
13. Dubois, D., Prade, H.: Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. Fuzzy Sets and Systems 192, 3–24 (2012)
14. van den Berg, J., van den Bergh, W.M., Kaymak, U.: Probabilistic and statistical fuzzy set foundations of competitive exception learning. In: Proceedings of the Tenth IEEE International Conference on Fuzzy Systems, Melbourne, Australia, vol. 2, pp. 1035–1038 (2001)
15. Waltman, L., Kaymak, U., van den Berg, J.: Fuzzy histograms: A statistical analysis. In: EUSFLAT-LFA 2005 Joint 4th EUSFLAT 11th LFA Conference, pp. 605–610 (2005)
16. Kaymak, U., van den Bergh, W.M., van den Berg, J.: A fuzzy additive reasoning scheme for probabilistic Mamdani fuzzy systems. In: Proceedings of the 2003 IEEE International Conference on Fuzzy Systems, St. Louis, USA, vol. 1, pp. 331–336 (2003)

17. Xu, D., Kaymak, U.: Value-at-risk estimation by using probabilistic fuzzy systems. In: Proceedings of the 2008 World Congress on Computational Intelligence, Hong-Kong, pp. 2109–2116 (2008)
18. Almeida, R., Kaymak, U.: Probabilistic fuzzy systems in Value-at-Risk estimation. Intelligent Systems in Accounting, Finance & Management 16, 49–70 (2009)
19. van den Berg, J., Kaymak, U., van den Bergh, W.M.: Financial markets analysis by using a probabilistic fuzzy modelling approach. International Journal of Approximate Reasoning 35, 291–305 (2004)
20. Kosko, B.: Fuzzy Engineering. Prentice Hall, Upper Saddle River (1997)
21. Kosko, B.: Global stability of generalized additive fuzzy systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 28, 441–452 (1998)
22. de Finetti, B.: Sull'impostazione assiomatica del calcolo delle probabilità. Annali Univ. Trieste 19, 3–55 (1949); (English translation: Probability, Induction, Statistics. Wiley, London, 1972 (Chapter 5)
23. Popper, K.: The Logic of Scientific Discovery. Routledge, London (1959)
24. Coletti, G., Scozzafava, R.: Conditional probability and fuzzy information. Computational Statistics & Data Analysis 51, 115–132 (2006)
25. Baldwin, J., Lawry, J., Martin, T.: A mass assignment theory of the probability of fuzzy events. Fuzzy Sets and Systems 83, 353–367 (1996)
26. Singpurwalla, N.D., Booker, J.M.: Membership functions and probability measures of fuzzy sets. Journal of the American Statistical Association 99, 867–877 (2004)

# Measure Inputs to Fuzzy Rules

Ronald R. Yager

Machine Intelligence Institute, Iona College
New Rochelle, NY
yager@panix.com

**Abstract.** Our concern is with the determination of the firing level of the antecedent fuzzy set in a fuzzy systems model rule base. We first consider the case where the input information is also expressed in terms of a normal fuzzy set. We provide the requirements needed by any formulation of this operation. We next consider the case when the input information is expressed using a measure. Here we also provide the requirements for any formulation that can be used to determine the firing level of the antecedent fuzzy set when the input information is a measure. We provide some examples of these formulations. Since a probability distribution is a special case of a measure we are able to determine the firing level of fuzzy rules with probabilistic inputs.

## 1 Introduction

Based on the ideas of Zadeh [1] fuzzy system modeling provides a method for building models of complex systems and mathematical functions [2, 3]. A fuzzy systems model consists of a rule base in which the antecedent conditions of the rules are expressed using normal fuzzy sets. Central to the use of these models is the determination of the firing level of the antecedent conditions based on the input to the model. In most cases the input to the fuzzy model has been expressed also using fuzzy sets. We provide a general formulation for the calculation of this value and look at possible implementations.

Another purpose here is to extend the capabilities of the fuzzy systems modeling technology by allowing a wider class of input information. Here we shall consider the case where the input is expressed using a measure representation [4, 5]. A notable special case is a probability distribution. In providing this extension a particularly important issue that arises is the determination of the firing level of a fuzzy set antecedent when the input information is expressed using a measure.

## 2 Determination of Individual Antecedent Firing Levels

In fuzzy systems modeling a typical rule is of the form

$$\text{If } V_1 \text{ is } A_1 \text{ and } V_2 \text{ is } A_2, \ldots \text{ and } V_n \text{ is } A_n \text{ then U is b,}$$

where the $A_j$ are normal fuzzy subsets over the domain of its associated variable the $V_j$. An important step in the use of these models is the determination of the firing

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 577–587, 2014.

levels for the individual antecedent components, the $V_j$ is $A_j$, given the knowledge about the value of the variable $V_j$, denoted, Inf(j). In the case where we know the exact value for $V_j$, $V_j = a_j$ then we simply get for $\tau_j$, the firing level of the component $V_j$ is $A_j$, that $\tau_j = A_j(a_j)$, the membership grade of $a_j$ in $A_j$.

Once we introduce some uncertainty or imprecision in our knowledge of the value of $V_j$ the situation becomes more complex. Let us consider the case where our knowledge is expressed by a normal fuzzy set $E_j$. One commonly used approach in this situation is to use the measure of possibility [6], $\tau_j = \text{Poss}[A_j/E_j] = \text{Max}_x[A_j(x) \wedge E_j(x)]$. One case that deserves attention here is where $E_j$ corresponds to no knowledge, $E_j = X_j$. In this case $\tau_j = 1$. Thus we get complete satisfaction of the requirement that $V_j$ is $A_j$ in the case when we don't know anything about the current value of $V_j$. This is somewhat troubling.

A second approach is to use the measure of necessity (certainty) [7]. In this case $\tau_j = \text{Cert}[A_j/E_j] = 1 - \text{Poss}[\overline{A_j}/E_j] = 1 - \text{Max}_x[\overline{A_j}(x) \wedge E_j(x)]$. Here we are determining the firing level as the negation of the possibility of not A being satisfied. Here we see here that if $V_j$ is precisely known, $E_j = \{a_j\}$ we still get $\tau_j = A_j(a_j)$. At the other extreme if we have no knowledge of $E_j(x) = 1$ for all x. Then $\tau_j = \text{Cert}[A_j/E_j] = 1 - \text{Max}_x[\overline{A_j}(x)] = 0$. Thus here we get a firing of the antecedent as zero. We note that if $E_j$ is normal then $\text{Cert}[A_j/E_j] \leq \text{Poss}[A_j/E_j]$.

The preceding two formulations are specific examples of a more general operation that determines the satisfaction of the requirement that $V_j$ is $A_j$ given $V_j = E_j$. We shall refer to this as the Validity of $A_j$ given $E_j$ and denote it as $\text{Val}(A_j/E_j)$. In the following, for notational simplicity we suppress the subscript j and simply use the notation Val(A/E). Here we are interested in satisfying the condition V is A given the knowledge V is E. In the following we provide a general characterization of the concept Val(A/E).

**Definition:** If A and E are two normal fuzzy subsets of X we say Val(A/E) is a quantification of the degree of satisfaction of A given E if Val(A/E) has the properties

1) If E is a singleton, E = {x), then Val(A/E) = A(x)
2) If A is the whole special, A = X, then for any E, Val(A/E) = 1
3) If $A \cap E = \varnothing$ the Val(A/E) = 0
4) Val(A/E) is monotonic in A, if $A_1 \subseteq A_2$ then $\text{Val}(A_1/E) \leq \text{Val}(A_2/E)$

We point out that the condition $A \cap E = \varnothing$ implies that $\text{Max}_x[A(x) \wedge E(x)] = 0$. We can show that the measures of possibility and certainty satisfy these conditions.

Another example of an operator that satisfies these conditions, is defined as $\text{Mid}(A/E) = \dfrac{\sum_j A(x_j)\, E(x_j)}{\sum_i E(x_i)}$. We can easily show that this is also a Val operator.

If E = X then Mid(A/E) = $\frac{\sum_i A(x_i)}{n}$. We see this as a compromise between the extremes of the Poss and Cert formulations which give one and zero respectively when E = X.

We now provide another formulation, CP(A/E), which we can show is a Val function. Here we let ρ be an index function so that ρ(k) is the index of the element in X with the $k^{th}$ largest value for $A(x_i)$, thus $A(x_{\rho(k)})$ is the $k^{th}$ largest membership grade in A. Using this we define

$$CP(A/E) = \sum_{j=1}^{n} (\underset{k \le j}{Max}[E(x_{\rho(k)})] - \underset{k \le j-1}{Max}[E(x_{\rho(k)})])\, A(x_{\rho(j)}).$$

In the above by convention for j = 1 we define $\underset{k \le j-1}{Max}[E(x_{\rho(k)})] = 0$. Using the notation $w_j = \underset{k \le j}{Max}[E(x_{\rho(k)})] - \underset{k \le j-1}{Max}[E(x_{\rho(k)})]$ we see $CP(A/E) = \sum_{j=1}^{n} w_j A(x_{\rho(j)})$.

Its clear that since both A and E are normal fuzzy sets all $w_j \ge 0$ and $\sum_{j=1}^{n} w_j = 1$ and hence CP(A/E) is a weighted average of the $A(x_i)$. We also note here that if E = X, then CP(A/E) = 1.

We now look at some general properties of validity operators.

**Theorem:** For any collection $Val_i(A/E)$, i = 1 to q, of validity operators and set of q weights $w_i \in [0, 1]$ and $\Sigma_i w_i = 1$ then $Val(A/E) = \Sigma_i w_i\ Val_i(A/E)$ is a validity operator.

Thus we see the linear combination of validity operators is a validity operator. A notable example of the use of the linear combination of validity operators is

$$Val(A/E) = \alpha Poss(A/E) + \bar{\alpha}\, Cert(A/E)$$

Another interesting example of the combination of these operators is

$$Val(A/E) = \alpha Poss(A/E) + \beta Cert(A/E) + (1 - (\alpha + \beta))Mid(A)$$

where we have $\alpha + \beta \le 1$.

We easily can show that if $Val_1$, …, $Val_q$ are validity operators then $Val(A/E) = Min_i[Val_i(A/E)]$ is a validity operator. However we point out if we replace Min by any other non-idempotent t–norm this does not hold. Also $Val(A/E) = Max_i[Val_i(A/E)]$ is a validity operator. We again note that this does not hold if we replace Max by another t-conorm S.

We observe if $Val_1$, … $Val_q$ are validity operators and H is any mean aggregation operator [8] then $Val(A/E) = H(Val_1(A/E), Val_2(A/E), …, Val_q(A/E))$ is a validity operator.

Consider now a generalization of possibility measure, $Poss_T(A/E) = Max_j[T(A(x_j), E(x_j))]$, here we replace Min by a t-norm T. We can show that $Poss_T(A/E)$ is a validity operator. More generally if T and S are t-norms and t-conorms and we define $Poss_{T, S}[A/E] = S_j[T(A(x_j), E(x_j))]$ then $Poss_{T, S}[A/E]$ is a validity operator. We also can show that if $Cert_{T, S}[A/E] = 1 - Poss_{T, S}[\bar{A}/E]$ then $Cert_{T, S}$ is a validity operator.

In the preceding we considered the situation where our knowledge about the variable V was in the form of fuzzy subset E. We now turn to the case where our knowledge about V is expressed using a measure.

## 3 Measure Representation of Uncertain Information

A fuzzy measure $\mu$ on space X is a mapping $\mu: 2^X \rightarrow [0, 1]$ such that $\mu(\varnothing) = 0$, $\mu(X) = 1$ and $\mu$ is monotonic, $\mu(A) \geq \mu(B)$ if $B \subseteq A$ [4, 9].

We note that for any fuzzy measure because of the monotonicity property we have $\mu(A \cup B) \geq Max[\mu(A), \mu(B)]$ and $\mu(A \cap B) \leq Min(\mu(A), \mu(B))$. Assume $\mu_1$ and $\mu_2$ are two measures such that for all A we have $\mu_1(A) \geq \mu_2(A)$ we denote this as $\mu_1 \geq \mu_2$. We can associate with any measure $\mu$ a dual measure $\hat{\mu}$ defined $\hat{\mu}(A) = 1 - \mu(\bar{A})$. It is clear that the dual of the dual is the original measure. Thus $\mu$ and $\hat{\mu}$ come in unique pairs.

Two notable examples of measure are $\mu^*$ and $\mu_*$. Here $\mu^*(A) = 1$ for $A \neq \varnothing$ and $\mu^*(\varnothing) = 0$ and $\mu_*(A) = 0$ for $A \neq X$ and $\mu(X) = 1$. These are duals of each other, $\hat{\mu}^* = \mu_*$ and $\hat{\mu}_* = \mu^*$. In addition $\mu^* \geq \mu_*$ for any measure $\mu$ we have $\mu_* \leq \mu \leq \mu^*$.

If V is a variable taking its value in the space X we can represent uncertain or imprecise information about the value of V using a measure $\mu$ on the space X. Under this representation the measure of a set A, $\mu(A)$, indicates our *anticipation* of finding the value of the variable V in A [5]. We see here that the dual $\hat{\mu}(A)$ is a measure of the anticipation of **not** finding the value of the variable in **not** A.

Let us look at the use of some special fuzzy measures using this perspective of representing information about an uncertain variable. If we know that V is exactly equal to *a* then we represent this by the measure $\mu_a$ such that $\mu_a(A) = 1$ if $a \in A$ and $\mu_a(A) = 0$ if $a \notin A$. We can show that $\hat{\mu}_a = \mu_a$. Here we say that $\mu$ is self-dual.

At the other extreme from the measure $\mu_a$ are the measure $\mu^*$ and $\mu_*$. These are used to represent situations in which we know nothing about the value of variables other than it that is lies in X. However these measures represent this ignorance in different ways. $\mu^*$ represents our ignorance in a very optimistic way, it anticipates that V lies in any set A, $\mu(A) = 1$, except $\varnothing$. $\mu_*$ represents our ignorance in a very

pessimistic way, it anticipates that V lies in no set except X. Thus while both $\mu_*$ and $\mu^*$ are representing ignorance they are doing it in different ways/

Another notable measure is $\mu_n(A) = \frac{|A|}{n}$. It is the number of elements in the set A divided by the total number of elements, n, in X. We see that $\mu_n(\{x_j\}) = 1/n$ for all $x_j$. While this is often used to represent situations when not knowing anything about the anticipation of the outcome, we note, however that this has the implicit assumption that all elements have the same anticipation. We also note that for this measure $\hat{\mu}_n = \mu$.

We can represent probabilistic uncertainty using a measure $\mu$ so that $\mu(A) = \sum_{x_i \in A} p_i$ where $p_i = \text{Prob}(x_i)$. Thus here $\mu(\{x_i\}) = p_i$. A very special case of probability measure is one in which $p_i = 1/n$ for all i, this is $\mu_n$. Another special case of probability is one in which $p_i = 1$ for one $x_i$ and $p_j = 0$ for $j \neq i$. This actually is the same as the case where we know the value of V is $x_i$. A probability measure is also self-dual.

Possibilistic uncertainty can also be represented in this framework using a possibility measure $\mu$. Here if $\pi_i$ is the possibility of $x_i$, then for any subset A, $\mu(A) = \underset{x_i \in A}{\text{Max}}[\pi_i]$. We easily see $\mu(\{x_i\}) = \pi_i$. We note that for possibility measure $\mu(A \cup B) = \text{Max}[\mu(A), \mu(B)]$ for $A \cap B = \varnothing$. An interesting special case of this is one in which $\pi_i = 1$ for $x_i \in A$ and $\pi_i = 0$ for all $x_i \notin A$. This can be seen as the possibility representation of the information that $V \in A$. Actually we see the measure $\mu^*$ is an example of possibility measure in which $\pi_i = 1$ for all i. This can be seen as a possibilistic representation of unknown.

Closely related to the possibility measure is the certainty measure. If $\mu$ is a possibility measure its dual is a certainty measure [4]. Thus with $\mu$ a possibility measure $\hat{\mu}(A) = 1 - \mu(\bar{A})$ is a certainty measure. It can be shown $\hat{\mu}(A \cap B) = \text{Min}(\hat{\mu}(A), \hat{\mu}(B))$. It can be shown that $\mu(A) \geq \hat{\mu}(A)$ and $\mu(A) + \hat{\mu}(A) \geq 1$.

Another special case of uncertainty is a cardinality-based uncertainty. Here for every subset A, $\mu(A) = \alpha_{|A|}$ where |A| is the cardinality of A. In this case $\alpha_i$ are a collection of weights so that $\alpha_i \in [0, 1]$ where $\alpha_i \leq \alpha_j$ if $j > i$. We also require that $\alpha_0 = 0$ and $\alpha_n = 1$. Two examples of cardinal measures are $\mu^*$ and $\mu_*$. For $\mu^*$ we have $\alpha_j = 1$ for $j > 0$ and $\alpha_0 = 0$. For $\mu_*$ we have we have $\alpha_j = 0$ for $j < n$ and $\alpha_n = 1$.

## 4 Firing Fuzzy Rules With Measure Information

Earlier we considered the problem of determining the firing level of the antecedent V is A given the knowledge that V is E where E is another fuzzy subset. We extend our capabilities to the case where our knowledge about V instead of being another fuzzy set is a measure μ. Our concern here is with the determination of Val(A/μ).

When our knowledge was V is E we required four conditions be satisfied by any formulation of Val(A/E). Two of these conditions involved only information about A, these were the requirement that if A = X we get Val(X/E) = 1 and the monotonicity condition $Val(A_1/E) \leq Val(A_2/E)$ if $A_1 \subseteq A_2$. Another of these conditions only involved information about E, this was the requirement that $Val(A/E) = A(x_i)$ is E is the knowledge that $V = x_i$, $E = \{x_i\}$. The other condition involved information about the relationship between A and E, the that if $A \cap E = \emptyset$ then Val(A/E) = 0.

The first two conditions can be easily generalized to the case where we have V is μ

1) If A = X then Val(A/μ) = 1
2) If $A_1 \subseteq A_2$ then $Val(A_1/\mu) \leq Val(A_2/\mu)$

The third condition is about case when the measure μ conveys the information that V assumes some specific value. This expressed as

3) if μ is a measure representing $V = x_i$, then $Val(A/\mu) = A(x_i)$

The fourth condition requires a little more thought before we can move it to case where we have V is μ. This condition said that if the information, V is E, is completely conflicting with the antecedent, V is A, then Val(A/E) = 0. When our knowledge is V is μ our fourth condition is that if μ and A are completely conflicting then Val(A/μ) = 0. We now turn to the definition of complete conflict between a measure μ and a fuzzy set A. We shall say a measure μ is completely conflicting with the antecedent V is A if for any $x_j$ having $A(x_j) \neq 0$ the measure μ has the property that $\mu(B \cup \{x_j\}) = \mu(B)$ for all subsets B of X. We see some immediate implications of the case when A and μ are conflicting. If $A(x_j) \neq 0$ then $\mu(\{x_j\}) = 0$. This follows since $\{x_j\} = \emptyset \cup \{x_j\}$ and $\mu(\emptyset) = 0$. From this we see that for any $x_k$ we have $Min(\mu(\{x_k\}), A(x_k)) = 0$, More generally, since Min is the largest t-norm, then for any t-norm T, $T(\mu(\{x_k\}), A(x_k)) = 0$ and in particular the product $\mu(\{x_k\})\, A(x_k) = 0$.

We now state another interesting property of a measure μ that is completely conflicting with a fuzzy set A. If μ is completely conflicting with the fuzzy subset A then for any subset B of X consisting of *only* elements that have A(x) > 0, then μ(B) = 0.

Furthermore we observe that if $A(x_j) > 0$ then $\mu(X - \{x_j\}) = 1$. This follows since $(X - \{x_j\}) \cup (x_j) = X$ and μ(X) = 1.

We shall now look for some formulations that allow us to provide a definition of Val(A/μ) that satisfies the four conditions. Let us now consider the possibility of using $Choq_\mu(A)$ to provide a formulation for Val(A/μ). In anticipation of this we shall say something about the Choquet integral [10-12].

Assume $\mu$ is a measure on a space $X = \{x_1, \ldots, x_n\}$ and let f be a function on X such that $f: X \rightarrow [0, 1]$. Thus for each $x_i \in X$ we have $f(x_i) \in [0, 1]$. The Choquet integral of the f with respect to the measure $\mu$ is defined as $Choq_\mu(f) = \sum_{j=1}^{n} (\mu(H_j) - \mu(H_{j-1}))$ $f(x_{\rho(j)})$ where $\rho(j)$ is the index of the element in X having the $j^{th}$ largest value for f. In addition let $H_j = \{x_{\rho(k)} | k = 1 \text{ to } j\}$, the subset of X having the j largest values for $f(x_i)$. If we let $w_j = \mu(H_j) - (H_{j-1})$ then $Choq_\mu(f) = \sum_{j=1}^{n} w_j \, f(\rho(j))$ where $w_j \geq 0$ and $\sum_{j=1}^{n} w_j = 1$, $Choq_\mu(f)$ is a mean aggregation of the $f(x_i)$.

We shall use the Choquet integral to provide a formulation for $Val(A/\mu)$. Here we use the membership function of A as the function f in $Choq_\mu(f)$, thus $Val(A/\mu) = Choq_\mu(A)$. We now show that this satisfies the four conditions we require of $Val(A/\mu)$.

1) We first see that if A = X then A(x) = 1 for all x and hence

$$Choq_\mu(A) = \sum_{j=1}^{n} w_j \, A(x_{\rho(j)}) = \sum_{j=1}^{n} w_j = 1$$

2) From the known monotonicity of the Choquet integral if $A_1 \subseteq A_2$, $A_1(x_i) \leq A_2(x_i)$, then $Choq_\mu(A_1) \leq Choq_\mu(A_2)$.

3) The case where $V = x_i$ corresponds to a measure, $\mu_i$ such that $\mu_i(B) = 1$ if $x_i \in B$ and $\mu_i(B) = 0$ if $x_i \notin B$. Here then $Choq_{\mu_i}(A) = \sum_{j=1}^{n} (\mu_i(H_j) - \mu_i(H_{j-1})) \, A(x_{\rho(j)})$, however we see that $\mu_i(H_j) - \mu_i(H_{j-1}) = 1$ when $x_i \in H_j$ and $x_i \notin H_{j-1}$. Thus occurs when $x_{\rho(j)} = x_i$ thus $Choq_\mu(A) = A(x_i)$.

4) Finally the requirement that if A and $\mu$ are conflicting then $Val(A/\mu) = 0$. As we earlier noted A and $\mu$ are completely conflicting if for any $x_j$ having $A(x_j) \neq 0$ the measure $\mu$ has the property that $\mu(B \cup \{x_j\}) = \mu(B)$ for all B. Consider now the Choquet integral in this case, $Val(A/\mu) = \sum_{j=1}^{n} (\mu(H_j) - \mu(H_{j-1})) \, A(x_{\rho(j)})$, where $\rho(j)$ is the index of the element with the $j^{th}$ largest value for $A(x_i)$. We see that the $x_{\rho(j)}$'s can be partitioned into two classes, j = 1 to K where $A(x_{\rho(j)}) > 0$ and j = K + 1 to n where $A(x_{\rho(j)}) = 0$. Here then $Val(A/\mu) = \sum_{j=1}^{K} (\mu(H_j) - \mu(H_{j-1})) \, A(x_{P(j)})$. Furthermore since $H_j = H_{j-1} \cup \{x_{P(j)}\}$ then $\mu(H_j) = \mu(H_{j-1} \cup \{x_{P(j)}\})$ for j = 1 to K. If A and $\mu$

are conflicting then $\mu(H_j) = \mu(H_{j-1})$ for all $x_{P(j)}$ where $A(x_{p(j)}) > 0$, in this case for all j = 1 to K. Thus if $A(x_{p(j)}) > 0$ then $\mu(H_j) - \mu(H_{j-1}) = 0$ and hence $Val(A/\mu) = 0$.

Thus we see that the use of Choquet integral, $Choq_\mu(A)$, satisfies all the conditions required of a formulation for $Val(A/\mu)$.

## 5 Validity for Various Types of Inputs

Using the formulation $Val(A/\mu) = Choq_\mu(A)$ let us look at the types of structures that are obtained for various different kinds information about V.

We have already shown that if our knowledge is $V = x_i$, then $Val(A/\mu_i) = A(x_i)$. Consider the case where $\mu$ is a probability measure. In this case $\mu(\{x_i\}) = p_i$ and for any crisp subset B, $\mu(B) = \sum_{x_i \in B} p_i$. Here $Val(A/\mu) = \sum_{j=1}^{n} (\mu(H_j) - \mu(H_{j-1}))A(x_{\rho(j)})$. Since $H_j = H_{j-1} \cup \{x_{\rho(j)}\}$ and $\mu$ is an additive measure then $\mu(H_j) - \mu(H_{j-1}) = \mu(\{x_{\rho(j)}\}) = p_{\rho(j)}$. Thus in the case of a probabilistic input $\mu$ we get $Val(A/\mu) = \sum_{j=1}^{n} p_j A(x_j)$.

This is essentially the probability of A under the probability measure $\mu$. This is Zadeh's definition of the probability of a fuzzy set [6].

We now consider the situation in which we have a cardinality-based measure as our source of information. A recall a cardinality-based measure is defined in terms of a collection of parameters $0 = h_0 \leq h_1 \leq h_2 \leq \ldots \leq h_n = 1$ such that then for any crisp subset B, $\mu(B) = h_{|B|}$. If we let $r_j = h_j - h_{j-1}$ for j = 1 to n, then we have $r_j \in [0, 1]$. Since $|H_j| = j$ then $\mu(H_j) = h_j$ and hence $\mu(H_j) - \mu(H_{j-1}) = r_j$ and therefore we get

$$Val(A/\mu) = Choq_\mu(A) = \sum_{j=1}^{n} (\mu(H_j) - \mu(H_{j-1}))A(x_{\rho(j)}) = \sum_{j=1}^{n} r_j (A_{\rho(j)})$$

This is an OWA aggregation [13] of the $A(x_i)$ with OWA weights $w_j = r_j$. For the two forms of unknown, $\mu^*$ and $\mu_*$, we respectively get $Val(A/\mu^*) = 1$ and $Val(A/\mu_*) = A_{\rho(n)}$, the minimal membership in A.

In the case where the input is of the form of a possibility measure were $\mu(\{x_i\}) = \pi_i$ and for any crisp subset B, $\mu(B) = \max_{x_i \in B}[\pi_i]$. We can show that $Val(A/\mu) = \sum_{j=1}^{n} (\max_{k=1 \text{ to } j}[\pi_{\rho(k)}] - \max_{k=1 \text{ to } j-1}[\pi_{\rho(k)}])A(x_{\rho(j)})$ where $\rho(j)$ is the index of the element with the $j^{th}$ largest membership grade in A. We note this is what we earlier referred to as CP type of formula. Thus here we get one of forms for determining the firing

level of the antecedent fuzzy subset A given as the input another fuzzy set E where $E(_i) = \pi_i$..

Another class of measures, closely related to possibility measures, are the necessity measures [7]. Necessity measures are defined by the property that $\mu(A \cap B) = \text{Min}[\mu(A), \mu(B)]$. It is well known that necessity measures are duals of possibility measures. If $\tilde{\mu}$ is a possibility measure the measure $\mu$ defined so that $\mu(A) = 1 - \tilde{\mu}(\bar{A})$ is a necessity measure. We see that if $\tilde{\mu}(\{x_i\}) = \pi_i$ then we have for the associated necessity measure that $\mu(X - \{x_i\}) = 1 - \pi_i = \beta_i$. We can define a necessity measure $\mu$ using the collection $\beta_1, \ldots, \beta_n$ of parameters so that $\beta_i \in [0, 1]$ and $\text{Min}_i[\beta_i] = 0$, at least one $\beta_i = 0$. Using these we have that for any crisp set A that $\mu(A) = \underset{x_i \notin A}{\text{Min}}[\beta_i]$.

Here we see that in the case of a necessity measure $\mu(H_j) = \underset{x_i \notin H_j}{\text{Min}}[\beta_i] = \underset{k>j}{\text{Min}}[\beta_{\rho(k)}]$. In addition $\mu(H_{j-1}) = \underset{x_i \notin H_{j-1}}{\text{Min}}[\beta_i] = \mu(H_j) \wedge \beta_{\rho(j)} = \underset{k > j-1}{\text{Min}}[\beta_{\rho(k)}]$. Using this in $\text{Val}(A/\mu) = \sum_{j=1}^{n}(\mu(H_j) - \mu(H_{j-1}))A(x_{\rho(j)})$ we get $\text{Val}(A/\mu) = \sum_{j=1}^{n}(\underset{k>j}{\text{Min}}[\beta_{\rho(k)}] - \underset{k > j-1}{\text{Min}}[\beta_{\rho(k)}])A(x_{\rho(j)})$.

If we denote $\beta_i = 1 - \pi_i$ then $\underset{k>j}{\text{Min}}[\beta_{\rho(k)}] = 1 - \underset{k>j}{\text{Max}}[\pi_{\rho(k)}]$ and $\underset{k>j-1}{\text{Min}}[\beta_{\rho(k)}] = 1 - \underset{k>j-1}{\text{Max}}[\pi_{\rho(k)}]$. Using this we get

$$w_j = \mu(H_j) - \mu(H_{j-1}) = \underset{k>j-1}{\text{Max}}[\pi_{\rho(k)}] - \underset{k>j}{\text{Max}}[\pi_{\rho(k)}]$$

Another class of measures is the Sugeno measures [9]. For this measure if $A \cap B = \varnothing$ then $\mu_\lambda(A \cup B) = \mu_\lambda(A) + \mu_\lambda(B) + \lambda\mu_\lambda(A)\mu_\lambda(B)$ where $\lambda > -1$. If $\mu_\lambda(\{x_i\}) = L_i$ then it can be shown that $\mu_\lambda(B) = \frac{1}{\lambda}[\prod_{x_i \in B}(1+\lambda L_i) - 1]$. Let us now calculate $\text{Val}(A/\mu_\lambda) = \text{Choq}_{\mu_\lambda}(A) = \sum_{j=1}^{n}(\mu_\lambda(H_j) - \mu_\lambda(H_{j-1}))A(x_{\rho(j)})$. We first observe that since $H_j = H_{j-1} \cup \{x_{\rho(j)}\}$ and $H_{j-1} \cap \{x_{\rho(j)}\} = \varnothing$ then

$$\mu_\lambda(H_j) = \mu_\lambda(\{x_{\rho(j)}\}) + \mu_\lambda(H_{j-1}) + \lambda\mu_\lambda(\{x_{\rho(j)}\})\mu_\lambda(H_{j-1})$$
$$\mu_\lambda(H_j) = L_{\rho(j)} + \mu_\lambda(H_{j-1}) + \lambda L_{\rho(j)}\mu_\lambda(H_{j-1})$$

We see that $\mu_\lambda(H_j) - \mu_\lambda(H_{j-1}) = L_{\rho(j)} + \lambda L_{\rho(j)}\mu_\lambda(H_{j-1}) = L_{\rho(j)}(1 + \lambda\mu_\lambda(H_{j-1}))$. Furthermore since $H_{j-1} = \{x_{\rho(k)}/k = 1 \text{ to } j-1\}$ then $\mu(H_{j-1}) = \frac{1}{\lambda}((\prod_{k \le j-1}(1+\lambda L_{\rho(k)}) - 1)$. Therefore

$$\mu_\lambda(H_j) - \mu_\lambda(H_{j-1}) = L_{\rho(j)}(1 + (\prod_{k \le j-1}(1+\lambda L_{\rho(k)})) - 1).$$

Hence we get the formulation that $w_j = L_{\rho(j)} \prod_{k \le j-1}(1+\lambda L_{\rho(k)})$. Using this form for $w_j$ we get

$$\mathrm{Val}(A/\mu_\lambda) = \sum_{j=1}^{n}(A(x_{\rho(j)})\, L_{\rho(j)} \prod_{x \le j-1}(1+\lambda L_{\rho(k)}))$$

We note if $\lambda = 0$ then we get the simple additive case and here $w_j = L_{\rho(j)}$.
A very clearly recursive relationship exists for generating the $w_j$

$$W_1 = L_{\rho(1)}$$

$$w_j = \frac{L_{\rho(j)}}{L_{\rho(j-1)}} \cdot (1+\lambda L_{\rho(j-1)})\, w_{j-1} \qquad \text{for } j > 1$$

## 6 Conclusion

Our focus was with the firing level of the antecedent fuzzy set in a fuzzy systems model. First we considered the case where the input was also expressed in terms of a normal fuzzy set. We provided the requirements for any formulation of this operation. We considered the case when the input information was a measure. We provided the requirements for any formulation that can be used to obtain the firing level of the antecedent fuzzy set when the input is a measure. We provided some examples of these formulations. Using these results we were able to determine the firing level of fuzzy rules with probabilistic inputs.

**Acknowledgement.** This work was supported by a ARO (MURI) grant (Number W911NF-09-1-0392) and by the ONR grant award number N00014-13-1-0626.

## References

[1] Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man, and Cybernetics 3, 28–44 (1973)
[2] Yager, R.R., Filev, D.P.: Essentials of Fuzzy Modeling and Control. John Wiley, New York (1994)
[3] Pedrycz, W., Gomide, F.: Fuzzy Systems Engineering: Toward Human-Centric Computing. John Wiley & Sons, New York (2007)

[4] Wang, Z., Klir, G.J.: Generalized Measure Theory. Springer, New York (2009)
[5] Yager, R.R.: A measure based approach to the fusion of possibilistic and probabilistic uncertainty. Fuzzy Optimization and Decision Making 10, 91–113 (2011)
[6] Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
[7] Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
[8] Beliakov, G., Pradera, A., Calvo, T.: Aggregation Functions: A Guide for Practitioners. Springer, Heidelberg (2007)
[9] Sugeno, M.: Fuzzy measures and fuzzy integrals: a survey. In: Gupta, M.M., Saridis, G.N., Gaines, B.R. (eds.) Fuzzy Automata and Decision Process, pp. 89–102. North-Holland Pub., Amsterdam (1977)
[10] Sugeno, M., Murofushi, T.: Choquet integral as an integral form for a class of fuzzy measures. In: Proceedings of the Second IFSA Congress, Tokyo, pp. 408–411 (1987)
[11] Yager, R.R.: Criteria aggregations functions using fuzzy measures and the Choquet integral. International Journal of Fuzzy Systems 1, 96–112 (1999)
[12] Klement, E.P., Mesiar, R., Pap, E.: A universal integral as common frame for Choquet and Sugeno. IEEE Transaction of Fuzzy Systems 18, 178–187 (2010)
[13] Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man and Cybernetics 18, 183–190 (1988)

# Interpretability in Fuzzy Systems Optimization: A Topological Approach

Ricardo de Aldama* and Michaël Aupetit

CEA, LIST, Laboratoire Analyse de Données et Intelligence des Systèmes,
Gif-sur-Yvette F-91191, France
ricardo.dealdama@csl.sony.fr, michael.aupetit@cea.fr

**Abstract.** When dealing with complex problems, it is often the case that fuzzy systems must undergo an optimization process. During this process, the preservation of interpretability is a major concern. Here we present a new mathematical framework to analyze the notion of interpretability of a fuzzy partition, and a generic algorithm to preserve it. This approach is rather flexible and it helps to highly automatize the optimization process. Some tools come from the field of algebraic topology.

**Keywords:** fuzzy system, fuzzy partition, optimization, tuning, interpretability, algebraic topology.

## 1 Introduction

One of the appealing features of fuzzy ruled based systems is that in most cases they are easily interpretable by humans. However, when used to tackle complex problems, there is often need to make use of automatic optimization methods that improve the original system (cf. [2]). These automatic methods have a drawback: It may entail important losses in the interpretability of the system, in particular in the fuzzy partitions. The goal of this paper is to deal with this loss of interpretability.

Although there is no standard definition for the notion of interpretability of a fuzzy system, we can distinguish, following [1,7], two levels of interpretability: That of fuzzy partitions and that of rule analysis. In this paper we deal with the problem of preserving the interpretability of the fuzzy partitions during the process of parametric optimization or tuning. We can divide this work in two parts: Firstly we provide a mathematical framework in which the concept of interpretability may be formalized, and secondly we provide a generic algorithm that takes as input a fuzzy system that the user considers interpretable, and a function to optimize (that measures the quality of a fuzzy system) and gives as output an optimized fuzzy system that preserves interpretability.

Thanks to this formalization the optimization process will be, in our view, much more painless for the user than in previous approaches. In particular it may be carried out not only by experts in optimization of fuzzy systems as usual,

* Current address: Sony CSL Paris, 6 rue Amyot, 75005 Paris, France.

A. Laurent et al. (Eds.): IPMU 2014, Part I, CCIS 442, pp. 588–597, 2014.

but also by users that are just experts in the problem-specific domain and whose knowledge in fuzzy theory may be limited.

In our approach we do not fix a priori the notion of interpretability. The mathematical framework that we propose is problem-independent and sufficiently generic to let the user establish which configuration he wants to preserve during the optimization. The essential point is the formalization of the notion of interpretability in topological and geometrical terms. Its preservation implies some particular constraints on the acceptable solutions for the optimization problem. In the generic algorithm that we propose, the codification and verification of these constraints is automatically done.

The geometric and topological analysis begins with a collection of fuzzy systems that the user considers interpretable (although in the our description of the algorithm we suppose that this family is composed of just one system). The domain of each variable is partitioned in such a way that the relative order of the different membership functions is constant on each region. These regions, and the order relations associated to them, will determine the geometric and topological constraints that will be taken into account during the optimization. In order to codify this information, a key role is played by homology groups. We make use of these well-known algebraic objects, which are able to capture a very significant part of the topology of a space and are well-suited for computer calculations. There exist several implementations to compute different homology groups. The reader interested in more details may consult for instance [4], [5] or [8].

## 2 Analyzing the Interpretability: A Topological Framework

### 2.1 The Main Idea

What we propose in this paper is not an absolute definition of interpretability, but rather a framework in which the actual definition, which will strongly depend on the user, can be expressed. We may talk then, given a user $U$, of *interpretability relative to* $U$. Our approach is strongly focused on topology: Our viewpoint is that the properties of the fuzzy partition that the user requires to be preserved are essentially of a topological nature.

Let us say a user defines a fuzzy partition such as the one on Figure 1 (top). It seems reasonable to consider that the user requires the optimization process to preserve, at least, the order of the terms. This order, although not explicitly formalized, underlies the solution we usually find in the literature: To strongly constrain the possible variations of the membership functions, in order to obtain very similar configurations as the original one, as in Figure 1 (top).

Some difficulties may arise if we try to define an order in a case such as that of Figure 1 (bottom). In more general cases, such as those of 2-dimensional variables, the concept of order may not even make any sense. However, there are always some basic properties that the user wants to preserve to be able to attach some meaning to the system. In our approach, these properties have

a topological nature and are locally determined by the values of the different membership functions. In particular, we think that the relative order of these values is crucial.

The main idea is to partition the numeric domain of the variable into regions in which the relative order of the membership functions is constant, such as in Figure 1 (bottom). We use the following notation: If $t_1, t_2, t_3$ are linguistic terms and $\mu_{t_1}, \mu_{t_2}, \mu_{t_3}$ their corresponding membership functions, then a region $R$ of the domain is said to satisfy the relation $t_1 > t_2 > t_3$, or have the label $t_1 > t_2 > t_3$, if for every $x \in X$ the relation $\mu_{t_1}(x) > \mu_{t_2}(x) > \mu_{t_3}(x)$ holds.

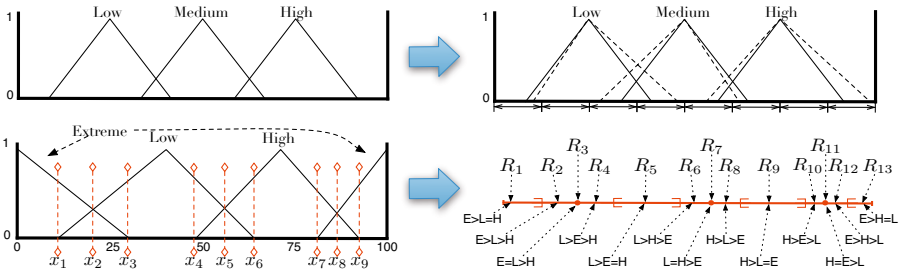



**Fig. 1.** Top: Example of a fuzzy partition and some typical constraints on it (classical approach). Bottom: Decomposition of the domain $[0, 100]$ in regions $R_i$ in which the relative order of the membership functions is constant (our approach). For instance, if $\mu_E$, $\mu_L$, $\mu_H$ are the membership functions corresponding to Extreme, Low, High, then $R_2$ gets the label Extreme > Low > High, since $\mu_E(x) > \mu_L(x) > \mu_H(x)\ \forall x \in R_2$.

Some properties of this partition will be required to be preserved during the optimization process. Examples of such properties could be:

- There is a region $R_2$ in which the relation Extreme > Low > High holds, with neighbors $R_1$ and $R_3$, such that in $R_1$ we have Low = High < Extreme, and in $R_3$ we have Extreme = Low > High.
- The value 50 belongs to the region $R_6$ that verifies Low > High > Extreme.

The rest of the section will be devoted to make this main idea more precise. In particular, we will present two key notions: The *geometric* and *topological signatures*.

### 2.2 Notation and Definitions

The definitions concerning fuzzy systems, such as linguistic variable, membership function, etc. are standard (see for instance [6]). We consider that the numeric domains associated to each linguistic variable are equipped with a natural topology (as it is the case with $\mathbb{R}^n$).

- Let $\Omega$ be the set of possible fuzzy systems under consideration, and let $A = A_1 \times \ldots \times A_n$ (typically $A \subseteq \mathbb{R}^n$) be the domain of the parameter vector that we consider as determining a fuzzy system. A solution to our optimization problem will be then an element $\bar{a} \in A$.
- We denote by $\omega : A \to \Omega$ the map that determines a fuzzy system $\omega(\bar{a})$ from the parameter vector $\bar{a}$. In particular $\omega$ determines every membership function of the system.
- We denote by $V$ the set of all linguistic variables and we suppose it is the same for every $\omega \in \Omega$. We denote by $Dom_v$ the domain of a linguistic variable $v \in V$.

### 2.3 Geometric Signature

Let $\omega \in \Omega$ be a fuzzy system and $v \in V$ a linguistic variable. The *geometric signature of $\omega$ relative to $v$*, that we denote by $\mathscr{G}_\omega(v)$, is a mathematical object that captures all the potentially interesting properties of the partition induced by $\omega$ on $Dom_v$. It provides the regions in which the relative order of the different membership functions is constant, and together with each region, its corresponding order.

As an illustration, consider that for a certain $\omega \in \Omega$ and $v \in V$, Figure 1 (bottom) represents the partition induced by $\omega$ on $Dom_v$. In this case $\mathscr{G}_\omega(v)$ is the map that associates to $i \in \{1, \ldots, 13\}$ the region $R_i$, together with the corresponding order relation on terms. For instance:

- $\mathscr{G}_\omega(v)(1)$ is the region $R_1$, i.e. the interval $[0, x_1]$, together with the label $\mathsf{Extreme} > \mathsf{Low} = \mathsf{High}$. This means that for every $x \in R_1$, $\mu_{\mathsf{Extreme}}(x) > \mu_{\mathsf{Low}}(x) = \mu_{\mathsf{High}}(x)$, where $\mu_{\mathsf{Extreme}}$, $\mu_{\mathsf{Low}}$, $\mu_{\mathsf{High}}$ are the corresponding membership functions.
- $\mathscr{G}_\omega(v)(3)$ is the region $R_3$, i.e. the set $\{x_2\}$ composed of only one point, together with the label $\mathsf{Extreme} = \mathsf{Low} > \mathsf{High}$. In practice, regions of low dimension (0 in this case) may be ignored.

In some cases the user might consider certain "dummy" functions $Dom_v \to [0, 1]$ to code particular constraints, such as interactions between membership functions. For instance, to deal with strong partitions we might consider the constant function 1 and the function $\sum_i \mu_i(x)$ (where $\mu_i$ represents the $i$-th membership function).

The *geometric signature of* $\omega$, denoted by $\mathscr{G}_\omega$, is the map that associates $\mathscr{G}_\omega(v)$ to $v \in V$.

### 2.4 Topological Signature

The *topological signature of $\omega$ relative to $v$*, that we denote by $\mathscr{T}_\omega(v)$, is a a weaker concept than that of the geometric signature, i.e. for $\omega, \eta \in \Omega$, if $\mathscr{G}_\omega(v) = \mathscr{G}_\eta(v)$ then $\mathscr{T}_\omega(v) = \mathscr{T}_\eta(v)$. It codes the topological information contained in $\mathscr{G}_\omega(v)$.

The *topological signature of* $\omega$ is the map that associates $\mathscr{G}_\omega(v)$ to $v \in V$. We denote by $\mathscr{T}_\omega$.

In the field of computational topology, the use of homology groups is widely spread to deal with the topology of a space. We will not provide here any definition concerning homology theory, since it is out of the scope of this paper; nevertheless we should say that these groups are topological invariants of algebraic nature, that capture an important part of the topological information of a space and are well-suited from an algorithmic viewpoint. The reader interested may consult for instance [4], a standard reference in algebraic topology, or [5] and [8] for an approach more focused on computational aspects.

We can propose then to code the topological signature in terms of these homology groups, that we denote by $H_N$ for $N \in \mathbb{N}$. Let $v \in V$ and consider $\omega, \eta \in \Omega$ such that $\omega$ induces a partition on $Dom_v$ composed of regions $R_1, \ldots, R_n$ and $\eta$ induces a partition on $Dom_v$ composed of regions $S_1, \ldots, S_n$. We say that $\mathscr{T}_\omega(v)$ and $\mathscr{T}_\eta(v)$ are equal if there is a $n$-permutation $\sigma$ such that:

1. the order on terms corresponding to $R_i$ is the same as that of $S_{\sigma(i)}$ for $i = 1, \ldots, n$, and moreover
2. $H_n(\bigcup_{k\in K} S_{h(k)}) \approx H_n(\bigcup_{k\in K} R_k)$ for each $K \subseteq I$ and $n \in \mathbb{N}$.

The homology groups are characterized by some integers, namely the *Betti numbers* and the *torsion coefficients*; they will be stored and used as topological signature. However, we should say that this is a general-purpose coding; in practice there may be different ways to implement the notion topological signature, depending mostly on the nature of $Dom_v$. In some cases the computation of these homology groups may not be necessary and a much more efficient coding can be devised.

To illustrate the notion of topological signature, consider that for a certain $\omega \in \Omega$ and $v \in V$, Figure 1 (bottom) represents the partition induced by $\omega$ on $Dom_v$. In this case, $\mathscr{T}_\omega(v)$ provides for each $i \in \{1, \ldots, 13\}$ the the order on terms corresponding to the region $R_i$, and for each $K \subseteq \{1, \ldots, 13\}$ the topological information of $\bigcup_{i\in K} R_i$. For instance, if we consider $K = \{4, 5\}$, $\mathscr{T}_\omega(v)$ codes the fact that $R_4 \cup R_5$ is connected, and if we consider $K = \{1, 6, 9\}$ the fact that $R_1 \cup R_6 \cup R_9$ is composed of three connected components. Essentially, $\mathscr{T}_\omega(v)$ codes the following information:

1. There are 13 regions $R_i$ (each one being a connected set),
2. the order on terms corresponding to $R_1$ is $\mathsf{Extreme} > \mathsf{Low} = \mathsf{High}$, that of $R_2$ is $\mathsf{Extreme} > \mathsf{Low} > \mathsf{High}$, etc.
3. $R_1$ is neighbor of $R_2$, $R_2$ is neighbor of $R_1$ and $R_3$, etc.

Hence if we consider another $\eta \in \Omega$ whose decomposition of $Dom_v$ is given by regions $S_1, \ldots, S_M$, then $\mathscr{T}_\eta(v) = \mathscr{T}_\omega(v)$ iff $M = 13$, and for some permutation $\sigma$ we have:

1. The order on terms corresponding to $S_{\sigma(1)}$ is $\mathsf{Extreme} > \mathsf{Low} = \mathsf{High}$, that of $S_{\sigma(2)}$ is $\mathsf{Extreme} > \mathsf{Low} > \mathsf{High}$, etc.
2. $S_{\sigma(1)}$ is neighbor of $S_{\sigma(2)}$, $S_{\sigma(2)}$ is neighbor of $S_{\sigma(1)}$and $S_{\sigma(3)}$, etc.

## 3 User Interactions: An Operational Definition of Interpretability

As we have already mentioned, we do not provide an absolute definition of interpretability, but rather, given a user $U$, a conceptual and operational framework to deal with interpretability relative to $U$. The goal of this section is to show how we can define and manipulate this interpretability relative to $U$, relaying on the notions presented in Section 2 and, importantly, on some interactions with $U$. We should mention that the interactions we present here seem to us flexible enough to cover most part of needs; however, other interactions could be consider. Our base hypothesis is that the notion of interpretability has essentially a topological flavor. An oversimplified version of this hypothesis would be :

**Assumption 1.** *For every user $U$, there is a family $\{\omega_1, \ldots, \omega_n\} = \Sigma \subset \Omega$ of representative systems, such that every $\eta \in \Omega$ considered as interpretable by $U$, satisfies $\mathscr{T}_\eta = \mathscr{T}_{\omega_i}$ for a certain $i \in \{1, \ldots, n\}$.*

Since we want to provide an operational definition of interpretability relative to $U$, we need, of course, some interaction with $U$. We suppose then that $\Sigma \neq \emptyset$ and that $U$ is capable of specifying $\Sigma$, i.e. providing a parameter vector $\bar{a}_i \in A$ such that $\omega(\bar{a}_i) = \omega_i$, for $i = 1, \ldots, n$. This first interaction, in which $U$ provides $\Sigma$, is the slightest interaction with $U$ that our method needs. However, if we want to make our method more flexible, we can allow $U$ to provide more information. Next we present the two other kind of interactions that we consider.

For the sake of simplicity, we suppose hereafter that $\Sigma$ is composed of only one system, $\Sigma = \{\omega_0\}$. The general case is easily deduced from this particular case (see the end of this section).

*Relaxation of the topological conditions.* This is basically a relaxation of Assumption 1. Once $U$ has provided $\Sigma = \{\omega_0\}$, one could consider that for a solution $\bar{a} \in A$ to be acceptable, i.e. such that $\omega(\bar{a})$ is interpretable relatively to $U$, $\bar{a}$ must satisfy $\mathscr{T}_{\omega(\bar{a})} = \mathscr{T}_{\omega_0}$. Instead, we may let the user relax this condition: He could omit, if he wishes, some of the topological conditions imposed by $\mathscr{T}_{\omega_0}$. Typically it may consist in merging different regions and requiring a relaxed order on terms; in this case the relaxed order should be compatible with the order of the merged regions (see example in Figure 3). This notion of compatibility could be easily formalized in terms of the lattice of partial orders on terms. This interaction with $U$ induces some topological conditions $C^t$ that a solution $\bar{a}$ must satisfy to be considered interpretable by $U$. For instance, if there is no relaxation, a solution $\bar{a}$ satisfies $C^t$ if and only if $\mathscr{T}_{\omega(\bar{a})} = \mathscr{T}_{\omega_0}$.

*Addition of geometric conditions.* $U$ may strengthen the conditions for a solution to be considered interpretable. This extra conditions are of a geometric rather than topological nature. This will allow $U$ to specify the regions to which certain points should belong. If we consider again Figure 1 (bottom), $U$ may want to include the condition "$0 \in R_1$", that is "0 should belong to the region indexed by

1", or more precisely "0 should belong to the region whose corresponding order on terms is Extreme > Low = High, that is neighbor of other region (namely $R_2$) whose corresponding order is Extreme > Low > High, that is neighbor of etc. ". It is clear that we can codify these kind of conditions in terms of the point 0 and the signature $\mathscr{T}_{\omega_0}$. We note by $C^g$ the geometric conditions imposed by this interaction with $U$.

*Definition of interpretability* The interactions we have just described allow us to provide the following definition of interpretablity: A solution is *interpretable relative to* $U$ if it satisfies the conditions $C^t$ and $C^g$. During the optimization process we will use this definition to test if a solution is valid or not. In the general case, in which $\Sigma$ is not necessarily composed by only one $\omega_0$, the definition is very similar. For each $\omega \in \Sigma$, the user goes through the same interactions: Firstly he can relax the topological conditions induced by $\mathscr{T}_\omega$, and secondly he can add geometric conditions related with $\mathscr{G}_\omega$; then these interactions provide conditions $C^t_\omega$ and $C^g_\omega$ for each $\omega \in \Sigma$. In this case, a solution $\bar{a}$ is *interpretable relative to* $U$, if there is $\omega \in \Sigma$ such that $\bar{a}$ satisfies the conditions $C^t_\omega$ and $C^g_\omega$.

## 4 Algorithm

We present here the different parts of a generic algorithm that fulfills our purpose: To optimize a given fuzzy system while preserving its interpretability. In Figure 2 we can see a scheme of this algorithm, but rather than explaining it in its more abstract form, we prefer to focus in the explanation of a particular example. The generic case will easily be induced from this description.

Let us consider a certain fuzzy system $\omega_0$ modeling a 2-dimensional problem, in which only one linguistic variable $v$ is involved. For instance there may be some rules involving the terms East, West and Center that are used to activate some procedures: We could imagine a fuzzy controller that produces policy decisions (e.g. public transports, taxes, etc.) for towns in a certain area, following rules of the type "If town $T$ is in region East then apply policy $P$ to $T$". An example of the membership functions associated to East, West and Center can be found in

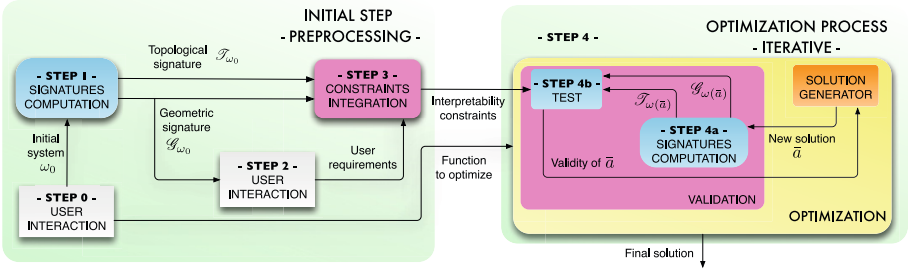



**Fig. 2.** Scheme of the algorithm. The steps correspond to the description in the text.

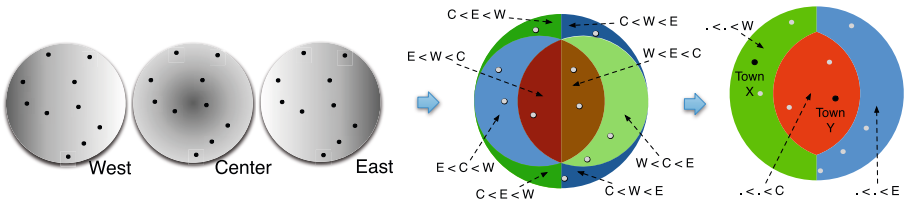



**Fig. 3.** Left: Example of three membership functions associated to a 2-dimensional variable (darker colors represent values closer to 1). The black dots represent towns. Center: Regions induced by the membership functions, with their corresponding labels. Right: A relaxation of the topological conditions and the addition of two geometric conditions: Since only the highest-valued functions are relevant, some labels are merged; moreover town $X$ must belong to the region in which $\mu_{\mathsf{West}}$ is the highest-valued function and town $Y$ to the region in which $\mu_{\mathsf{Center}}$ is the highest-valued function.

Figure 3 (left). Let us say a user $U$ considers $\omega_0$ as interpretable and wants to optimize it using a performance function $f$.

### Preprocessing

*Step 0.* The user gives $\omega_0$ and $f$ as input.

*Step 1.* The first part of the algorithm consists in computing the geometric signature, that is the regions in which the order of terms is constant. Let $\mu_{\mathsf{West}}, \mu_{\mathsf{Center}}, \mu_{\mathsf{East}}$ be the membership functions corresponding to the terms West, Center, East. The domain is discretize and each function is evaluated on each point of the grid. This evaluation induces a label for each point, e.g. a point $x$ gets the label $\mathsf{West} < \mathsf{East} < \mathsf{Center}$ if $\mu_{\mathsf{West}}(x) < \mu_{\mathsf{East}}(x) < \mu_{\mathsf{Center}}(x)$. Then we can explicitly compute the regions (maximal connected components with the same label) by using, for instance, the method described in [3]. See Figure 4 (top).

*Step 2.* At this point comes the second interaction with $U$ (apart from Step 0): The regions are presented to him (we can omit regions of dimension 0 and 1) and then he can, first relax the topological conditions that will be imposed to the acceptable (interpretable) solutions, and afterwards impose some geometric conditions. In Figure 3 we can see an example in which $U$, only interested in the function with highest value, decides to relax the topological conditions by merging the regions that share the same highest-valued function; he also imposes the geometric conditions "town $X$ must belong to the region in which the value of West is the biggest" and "town $Y$ must belong to the region in which the value of Center is the biggest".

*Step 3.* No other interaction with $U$ is needed, since he has just operationally defined what he considers as interpretable: This definition is essentially contained

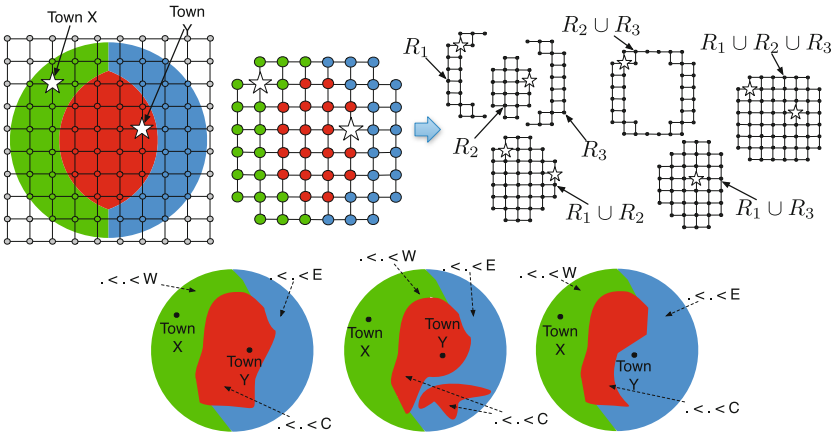



**Fig. 4.** Top: Signature computation (step 1: discretization and computation of regions). The Betti numbers of each region will be used to code the topology. Bottom left: A possible configuration of a solution that is acceptable. Bottom center: A configuration that does not satisfy the topological conditions since the region whose highest-valued function is $\mu_{\mathsf{Center}}$ is disconnected. Bottom right: A configuration that does not satisfy the geometric conditions since town $Y$ does not belong to the region whose highest-valued function is $\mu_{\mathsf{Center}}$.

in the right side of Figure 3. In Figure 4 (bottom) we can find examples of interpretable and not-interpretable solutions. This topological information is then coded in terms of homology groups, following the explanations of Section 2 and using for instance the algorithms presented in [5].

### Optimization Process

*Step 4.* This well-coded information, as well as the function $f$ and $\omega_0$, is given as an input to an optimization algorithm, and is interpreted as a constraint $C$ on the (signatures of the) solutions. This optimization algorithm may be of different types (e.g. metaheuristic or exact) depending on the nature of $f$. As it is the case for any iterative optimization algorithm, it should contain a "solution generator" module. This module may have different ways of dealing with constraints. The most basic option would be to test $C$ for each solution that it generates and to use the result of the test to generate a new solution. Another option would be to do some kind of preprocessing, in which the acceptable domain is approximated, and then to only generate valid solutions. In any case we will need to iterate a process similar to Step 1 and Step 3: Given $\bar{a} \in A$, compute $\mathscr{G}_{\omega(\bar{a})}$ and $\mathscr{T}_{\omega(\bar{a})}$, and use them to test if $\bar{a}$ satisfies $C$ (these are Step 4a and Step 4b in Figure 4). This will ensure that the final solution is interpretable relative to $U$.

## 5 Conclusion and Perspectives

We have presented a generic method to deal with the loss of interpretability in fuzzy partitions during the optimization of a fuzzy system. It relies essentially on topological concepts and tools, which confers a solid mathematical foundation. Our definition of interpretability is not absolute, but rather relative to each user, who implicitly defines the notion by means of some specific interactions. That makes this approach very flexible. Moreover, we think this method is sufficiently general to be uniformly applicable to most situations, without the need of an expert in optimization of fuzzy systems. This claim is justified by the fact that the user interactions are straightforward (no need of expertise in fuzzy systems optimization), as we can see in the description given in Section 4. This notion of interpretability could also open the door to other uses, independent of optimization; for instance, the quality of a system could be influenced by the complexity of its signature.

There are nevertheless some technical limitations. In the general case the computations are expensive, for instance that of homological groups. Also, if the membership functions are highly irregular, this might create lots of different regions to analyze. However, in a typical case of dimension 1 or 2 and simple functions (such as triangles) this should not be an issue. Other pathological cases may arise, such as a solution inducing a valid topology in which some regions are extremely small and not perceptible by the user. These limitations need some careful study, but do not seem impossible to overcome. We are currently working on the implementation of some case studies for user testing.

## References

1. Alonso, J.M., Guillaume, S., Magdalena, L.: Un indice d'interprétabilité des bases de règles floues. In: Rencontres LFA francophones sur la Logique Floue et ses Applications (LFA), Toulouse (2006)
2. Cordón, O., Herrera, F., Hoffmann, F., Magdalena, L.: Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases. World Scientific (2002)
3. Dillencourt, M., Samet, H., Tamminen, M.: A General Approach to Connected-Component Representations Labeling for Arbitrary Image Representation. Journal of the ACM (JACM) 39(2), 253–280 (1992)
4. Hatcher, A.: Algebraic topology. Cambridge University Press (2002)
5. Kaczynski, T., Mischaikow, K., Mrozek, M.: Computational homology. Applied Mathematical Sciences, vol. 157. Springer, New York (2004)
6. Klir, G.J., Yuan, B.: Fuzzy sets and fuzzy logic: theory and applications. Prentice Hall, New Jersey (1995)
7. de Oliveira, J.: Semantic constraints for membership function optimization. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 29(1), 128–138 (1999)
8. Zomorodian, A.: Topology for computing. In: Cambridge Monographs on Applied and Computational Mathematics, vol. 16, Cambridge University Press (2005)

# Author Index