Studies in Fuzziness and Soft Computing

David A. Pelta
Carlos Cruz Corona *Editors*

Soft Computing Based Optimization and Decision Models

To Commemorate the 65th Birthday of Professor José Luis "Curro" Verdegay

Springer

Studies in Fuzziness and Soft Computing

Volume 360

About this Series

The series "Studies in Fuzziness and Soft Computing" contains publications on various topics in the area of soft computing, which include fuzzy sets, rough sets, neural networks, evolutionary computation, probabilistic and evidential reasoning, multi-valued logic, and related fields. The publications within "Studies in Fuzziness and Soft Computing" are primarily monographs and edited volumes. They cover significant recent developments in the field, both of a foundational and applicable character. An important feature of the series is its short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

More information about this series at http://www.springer.com/series/2941

David A. Pelta · Carlos Cruz Corona
Editors

Soft Computing Based Optimization and Decision Models

To Commemorate the 65th Birthday of Professor José Luis "Curro" Verdegay

Springer

Editors
David A. Pelta
Department of Computer Science and A.I
Higher Technical School of Information Technology and Telecommunications Engineering
University of Granada
Granada
Spain

Carlos Cruz Corona
Department of Computer Science and A.I
Higher Technical School of Information Technology and Telecommunications Engineering
University of Granada
Granada
Spain

ISSN 1434-9922 ISSN 1860-0808 (electronic)
Studies in Fuzziness and Soft Computing
ISBN 978-3-319-64285-7 ISBN 978-3-319-64286-4 (eBook)
DOI 10.1007/978-3-319-64286-4

Library of Congress Control Number: 2017947475

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Jose Luis Verdegay, “Curro” for his friends, has greatly influenced our academic lives in the University of Granada, as much for his special dedication to our Department as Director, his successful management of the research, and especially as supervisor of our doctoral thesis.

He introduced us to the Fuzzy Sets Theory in the late 80’s, beginning of the 90’s and its application in optimization and decision making. He taught us to be committed to our University and scientific rigor. Under his supervision, we have learned to carry out serious research work. He has been a demanding advisor with the work that we have developed under his supervision, but he has always motivated us to focus on new challenges and, in such a way, we have grown scientifically and humanly. He has taught us to work as a team and, on the other hand, he has encouraged us to achieve scientific leadership in our academic lives.

As a result, we can say that we have got many scientific achievements and international recognition. Today we can assure that the Department of Computer Science and Artificial Intelligence is a reference team in our scientific community.

Special mention is made for his excellent activity in Latin America. He has been able to connect with many researchers from different universities in countries such as Cuba, Peru, Ecuador, etc., helping them in the development of their scientific careers.

His dedication and recognition for his initiatives at the many institutional positions he occupied, such as the most recent Rector’s Delegate of the University of Granada for Information and Communication Technologies, is very well known.

He is a national and international leader in our scientific community. He has the role of “Invited Professor” in three Cuban institutions (Instituto Superior Politécnico “José Antonio Echeverría” from La Habana, Universidad Central “Marta Abreu” from Las Villas and Universidad “Oscar Lucero Moya”, from Holguín). Also, he is a fellow of the International Fuzzy Systems Association (IFSA), he is recognized as an IEEE Senior Member, and he is a Honorary Member of the Mathematics and Computation Academy from Cuba.

Today more than ever, our friend Curro is still working actively in our Department and in our University.

For us, he represents the figure of our scientific and academic father with whom we share our scientific and academic achievements. It is always a pleasure to enjoy a good meal and cup of wine with him in one of our favorite restaurants “Chikito” or “Antonio Pérez”.

We want to use these brief words to thank you for everything you have done for us and to acknowledge your laudable dedication to the Department of Computer Science and Artificial Intelligence and the University of Granada.

Thank you very much CURRO!!!

Francisco “Paco” Herrera
Enrique Herrera-Viedma
Department of Computer Science and Artificial Intelligence
University of Granada
Spain

Preface

Decision and optimization problems are ubiquitous in the current social and technological context. Our societies face several challenges (in health, transportation, energy, climate, etc.) which are clearly framed either as decision or optimization problems. Moreover, research on Soft Computing is a key aspect in paving the way for better models and tools to solve the corresponding problems.

This commemorative book titled *Soft Computing based Optimization and Decision Models. To commemorate the 65th birthday of Professor José Luis "Curro" Verdegay*, contains 18 guest chapters addressing hot topics on soft computing based decision and optimization models and tools. They are written by key leading experts in the field from the USA, Brazil, UK, France, Cuba, Finland, Italy, Spain, etc., in which the reader will find short surveys, theoretical research and practical applications on the latest advances in the field.

The book is organized into three parts.

The first one comprises five chapters that review the main applications of Soft Computing in different fields.

In Chapter "A Review of Soft Computing Techniques in Maritime Logistics and Its Related Fields", Expósito-Izquierdo et al. highlight the role and relevance of maritime logistics and associated problems, and review the applications of soft computing techniques in the field. Opportunities for further developments are also explored.

J. Cadenas and M. Garrido, in Chapter "Intelligent Data Analysis, Soft Computing and Imperfect Data", analyze different hybridization approaches between soft computing and intelligent data analysis, focusing on the data preprocessing and data mining stages. They mainly focus on evaluating whether the elements of soft computing are incorporated in the design of the method/model, or whether they are also used to deal with imperfect information.

Chapter "Soft Computing Methods in Transport and Logistics" by J. Brito et al. begins by providing an overview of transport and logistic problems and their models focusing on the management of uncertainty by means of fuzzy optimization and metaheuristics methods. Then, and given the promising results, some emerging areas are presented and described.

Masegosa et al. contribute with the Chapter "Applications of Soft Computing in Intelligent Transportation Systems". Intelligent transportation systems combine electronic, communication and information technologies with traffic engineering to obtain more efficient, reliable and safer transportation systems. The chapter gathers and discusses some of the most relevant and recent advances in the application of soft computing techniques in relevant areas of intelligent transportation systems, namely autonomous driving, traffic state prediction, vehicle route planning and vehicular ad hoc networks.

Finally, in Chapter "Fuzzy Cognitive Maps Based Models for Pattern Classification: Advances and Challenges", G. Napoles et al. focus on Fuzzy Cognitive Maps (FCMs), a sort of recurrent neural networks that include elements of fuzzy logic during the knowledge engineering phase. The authors observe that many studies show how this soft computing technique (FCM) is able to model complex and dynamic systems, but here, they explore a new approach: the use of FCMs in solving pattern classification problems.

The second part of the book contains six contributions.

The first one is Chapter "A Proposal of On-Line Detection of New Faults and Automatic Learning in Fault Diagnosis", by A. Rodríguez Ramos et al. The authors present a new approach to automatic learning for a fault diagnosis system. The proposal includes an off-line learning stage, fuzzy clustering techniques and a metaheuristic (differential evolution). Then a novel fault detection algorithm is applied. This algorithm is able to determine whether an observation may constitute a new class, probably representative of a new fault or whether it is noise. The approach is validated using an illustrative example.

Then, two chapters deal with the portfolio selection problem. In the first one, Chapter "Fuzzy Portfolio Selection Models for Dealing with Investor's Preferences", C. Calvo et al. recall their previous works and propose a fuzzy model for dealing with the vagueness of investor preferences on the expected return and the assumed risk, and then consider several modifications to include additional constraints and goals. In the second one, Chapter "On Fuzzy Convex Optimization to Portfolio Selection Problem", R. Coelho departs from the fact that the portfolio selection problems can be classified as convex programming problems. Then, he presents a fuzzy set based method that solves a class of convex programming problems with vagueness costs in the objective functions and/or order relation in the set of constraints. The solution approach transforms a convex programming problem under fuzzy environment into a parametric convex multi-objective programming problem. The method is applied to a portfolio selection problem using the data of some Brazilian securities.

C. Carlsson, in Chapter "Digital Coaching for Real Options Support", claims that classical management science is making the transition to analytics and that there is a growing interest in replacing the classical net present value (NPV) with real options theory, especially for strategic issues and uncertain, dynamic environments. Both factors motivate the use of soft computing. As real options theory requires rather advanced levels of analytics, the author suggests that digital coaching is a way to guide and support users in giving them better chances for effective and productive use of real options methods. A real-world example on the development and use of fuzzy real options models for the case of closing (or not closing or closing later) an old paper mill in the UK is shown.

Chapter "An Analysis of Decision Criteria for the Selection of Military Training Aircrafts", by J. Sanchez Lozano et al. also fits in the context of decision making. The authors describe the process by which the relevance of technical criteria in determining the quality of a military training aircraft is obtained. Experts provided the criteria information and both qualitative and quantitative criteria are considered. A fuzzy AHP (Analytic Hierarchy Process) methodology is proposed to extract the knowledge from the group of experts and finally obtain a unique set of weights for the criteria.

Y. Liu and F. Gomide, in Chapter "Participatory Search in Evolutionary Fuzzy Modeling", focus on one of the key elements of soft computing, namely meta-heuristics. They introduce the so-called participatory search, a class of population-based search algorithms constructed upon the participatory learning paradigm. To illustrate the potential of the proposal, they resort to the problem of obtaining fuzzy rule-based models from actual data and provide comparisons with a state-of-the-art genetic fuzzy system.

The third part contains seven chapters exploring theoretical aspects of soft computing.

With the Chapter titled "But, What is It Actually a Fuzzy Set?", E. Trillas states that the idea of a fuzzy set is not yet clear enough and discusses the concept of fuzzy set as a quantity in whatever universe of discourse, and its possible use in the context of 'Computing with Words'.

D. Dubois and H. Prade contributed with the Chapter "Gradual Numbers and Fuzzy Solutions to Fuzzy Optimization Problems". The authors start with the idea of fuzzy elements in a fuzzy set, that is, entities that assign elements to membership values, in contrast with fuzzy sets that assign membership values to elements. Then, establishing a clear connection with the work of J.L. Verdegay, they observe that the fuzzy solution to a fuzzy optimization problem is a very early example of a fuzzy element in (or a gradual subset of) the fuzzy constraint set.

R. Yager, in Chapter "Using Fuzzy Measures to Construct Multi-criteria Decision Functions" explores the formulation of multi-criteria decision functions based on the use of a measure over the space of criteria, where the relationship among the criteria is expressed using a fuzzy measure. Such a fuzzy measure is used within the Choquet integral to construct decision functions and several specific cases are outlined.

Chapter "A Modal Account of Preference in a Fuzzy Setting", by F. Esteva et al. considers the problem of extending fuzzy preference relations on a set, to fuzzy preferences on subsets, and characterize different possibilities. They then propose several two-tiered graded modal logics to reason about the corresponding different notions of fuzzy preferences.

R. Fuller and I.Á. Harmati, in Chapter "On Possibilistic Dependencies: A Short Survey of Recent Developments", present a survey of the latest work on the extensions and developments of the notions of possibilistic mean value and variance of fuzzy numbers, possibilistic covariance, correlation ratio and correlation coefficient and the informational coefficient of correlation.

In Chapter "Penalty Function in Optimization Problems: A Review of Recent Developments", H. Bustince et al. highlight the role and relevance of penalty functions as a tool for information fusion. They review the ideas of penalty and penalty-based functions and discuss how such notions can be extended to deal with data in Cartesian products of lattices.

Finally, S. Bortot et al. in Chapter "The Single Parameter Family of Gini Bonferroni Welfare Functions and the Binomial Decomposition, Transfer Sensitivity and Positional Transfer Sensitivity" analyze the so-called generalized Gini welfare functions and consider their binomial decomposition. They introduce measures of transfer sensitivity and positional transfer sensitivity and illustrate the behaviour of the binomial welfare function with respect to these measures.

As Editors, we should highlight that it was both a challenge and a great pleasure for us in compiling this book.

On the one hand, it was a challenge because Prof. Verdegay has many friends and colleagues worldwide and we needed to select some of them as potential collaborators. The task was difficult but, in the end, we have an excellent set of topics written by top researchers, who collaborate or have collaborated with Curro. Here we thank the researchers who immediately accepted to join this editorial project. Readers of this book will appreciate these high-quality contributions.

In addition, we thank Profs. Enrique Herrera and Francisco Herrera for writing the foreword of the book.

On the other hand, it was a pleasure because we have known Curro since a long time ago. We started to work with him in 1998 (D. Pelta) and 2002 (C. Cruz) when we began our Ph.D. studies under his direction. From that time, we have had the opportunity to share many discussions, talks and personal situations with Curro that make us consider him a true friend. During these years, we have come to know all of Curro's facets. His scientific and academic merits are very well known, but we would like to mention here also his kindness, availability and true support for the academics and friends mainly from developing countries (especially Latin America). Since his work in the University Government, we have observed the huge number of visits he receives at his office asking for guidance or suggestions. His experience as a researcher, professor and manager (in several positions at the University Government) is invaluable and we are lucky to have him available every day.

Thank you Curro!

Acknowledgements D. Pelta and C. Cruz acknowledge the support of projects TIN2014-55024-P (Spanish Ministry of Economy and Competitiveness) and P11-TIC-8001 (Consejería de Economía, Innovación y Ciencia, Junta de Andalucía). Both projects include FEDER funds from the European Union.

Granada, Spain
May, 2017

David A. Pelta
Carlos Cruz Corona

Contents

Contributors

Rafael Bello Central University of Las Villas, Santa Clara, Cuba

Silvia Bortot Department of Economics and Management, University of Trento, Trento, Italy

Julio Brito Grupo de Computación Inteligente, Instituto Universitario de Desarrollo Regional, Universidad de La Laguna, La Laguna, Spain

Pedro Burillo Departamento de Automatica y Computacion, Universidad Publica de Navarra, Pamplona, Spain

Humberto Bustince Departamento de Automatica y Computacion, Universidad Publica de Navarra, Pamplona, Spain

Jose M. Cadenas Department of Information and Communications Engineering, University of Murcia, Murcia, Spain

Clara Calvo Universidad de Valencia, Valencia, Spain

Christer Carlsson IAMSR/Abo Akademi University, Turku, Finland

Dagoberto Castellanos-Nieves Grupo de Computación Inteligente, Instituto Universitario de Desarrollo Regional, Universidad de La Laguna, La Laguna, Spain

Ricardo Coelho Departamento de Estatística e Matemática Aplicada, Centro de Ciências, Universidade Federal do Ceará, Fortaleza, Brazil

Didier Dubois CNRS-IRIT, Université Paul, Toulouse, Cedex 9, France

Francesc Esteva IIIA - CSIC, Bellaterra, Spain

Christopher Expósito-Izquierdo Universidad de La Laguna, Santa Cruz de Tenerife, Spain

Airam Expósito Grupo de Computación Inteligente, Instituto Universitario de Desarrollo Regional, Universidad de La Laguna, La Laguna, Spain

Mario Fedrizzi Department of Industrial Engineering, University of Trento, Trento, Italy

Javier Fernandez Departamento de Automatica y Computacion, Universidad Publica de Navarra, Pamplona, Spain

Robert Fullér Department of Informatics, Széchenyi István University, Győr, Hungary

M.A. Socorro García-Cascales Dpto. de Electrónica, Tecnología de Computadoras y Proyectos, Universidad Politécnica de Cartagena, Murcia, Spain

M. Carmen Garrido Department of Information and Communications Engineering, University of Murcia, Murcia, Spain

Lluís Godo IIIA - CSIC, Bellaterra, Spain

Fernando Gomide School of Electrical and Computer Engineering, University of Campinas, Sao Paulo, Brazil

Isel Grau Central University of Las Villas, Santa Clara, Cuba

István Á. Harmati Department of Mathematics and Computational Sciences, Széchenyi István University, Győr, Hungary

Carlos Ivorra Universidad de Valencia, Valencia, Spain

María T. Lamata Dpto. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Granada, Spain

Maikel Leon Espinosa University of Miami, Miami, FL, USA

Pedro Lopez-Garcia Faculty of Engineering, University of Deusto, Bilbao, Spain; DeustoTech-Fundacion Deusto, Deusto Foundation, Bilbao, Spain

Vicente Liern Universidad de Valencia, Valencia, Spain

Yi Ling Liu School of Electrical and Computer Engineering, University of Campinas, Sao Paulo, Brazil

Orestes Llanes-Santiago Departamento de Automática y Computación, Universidad Tecnológica de la Habana José Antonio Echeverría, CUJAE, La Habana, Cuba

Ricardo Alberto Marques Pereira Department of Economics and Management, University of Trento, Trento, Italy

Antonio D. Masegosa Faculty of Engineering, University of Deusto, Bilbao, Spain; DeustoTech-Fundacion Deusto, Deusto Foundation, Bilbao, Spain; IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Belén Melián-Batista Universidad de La Laguna, Santa Cruz de Tenerife, Spain

J. Marcos Moreno-Vega Universidad de La Laguna, Santa Cruz de Tenerife, Spain

José. A. Moreno Grupo de Computación Inteligente, Instituto Universitario de Desarrollo Regional, Universidad de La Laguna, La Laguna, Spain

Enrique Onieva Faculty of Engineering, University of Deusto, Bilbao, Spain; DeustoTech-Fundacion Deusto, Deusto Foundation, Bilbao, Spain

Eneko Osaba Faculty of Engineering, University of Deusto, Bilbao, Spain; DeustoTech-Fundacion Deusto, Deusto Foundation, Bilbao, Spain

Gonzalo Nápoles Hasselt Universiteit, Diepenbeek, Belgium

Henri Prade CNRS-IRIT, Université Paul, Toulouse, Cedex 9, France

Alberto Prieto Moreno Departamento de Automática y Computación, Universidad Tecnológica de la Habana José Antonio Echeverría, CUJAE, La Habana, Cuba

Adrián Rodríguez Ramos Departamento de Automática y Computación, Universidad Tecnológica de la Habana José Antonio Echeverría, CUJAE, La Habana, Cuba

Anastasia Stamatopoulou Department of Economics and Management, University of Trento, Trento, Italy

Juan M. Sánchez-Lozano Centro Universitario de la Defensa. Academia General del Aire. Universidad Politécnica de Cartagena, Murcia, Spain

Enric Trillas University of Oviedo, Asturias, Spain

Koen Vanhoof Hasselt Universiteit, Diepenbeek, Belgium

Amanda Vidal ICS, Czech Academy of Sciences, Prague, Czech Republic

Ronald R. Yager Machine Intelligence Institute, Iona College, New Rochelle, NY, USA

Antônio José da Silva Neto Instituto Politécnico da Universidade Do Estado Do Rio de Janeiro (IPRJ/UERJ), Nova Friburgo, Brazil

A Review of Soft Computing Techniques in Maritime Logistics and Its Related Fields

Christopher Expósito-Izquierdo, Belén Melián-Batista and J. Marcos Moreno-Vega

Abstract The incessant increase in the world seaborne trade over the last few decades has encouraged maritime logistics has become a very attractive area of study for applying the general frameworks of soft computing. In this environment, there is a significant lack of efficient approaches aimed at obtaining exact solutions of a wide variety of optimization problems arisen in this field and which are classified as hard from the perspective of the complexity theory. These optimization problems demand increasingly new computational approaches able to report inexact solutions by exploiting extensively uncertainty, tolerance for imprecision, and partial truth to achieve tractability, among others. In the chapter at hand, we provide a review of the most highlighted soft computing techniques implemented in maritime logistics and its related fields and identify some opportunities to go further into depth on knowledge.

Keywords Optimization · Maritime container terminal · Logistics

1 Introduction

Over the last few decades, maritime container terminals have become outstanding infrastructures in global supply chains [35]. They are usually situated within the boundaries of ports located in strategic regions with the aim of acting as economic engines. The main purpose of these infrastructures is to carry out the efficient exchange of freights among heterogeneous means of transportation. Traditionally, these means have different operational and technical characteristics. Firstly, we can

C. Expósito-Izquierdo (✉) · B. Melián-Batista · J.M. Moreno-Vega
Universidad de La Laguna, Santa Cruz de Tenerife, Spain
e-mail: cexposit@ull.edu.es

B. Melián-Batista
e-mail: mbmelian@ull.edu.es

J.M. Moreno-Vega
e-mail: jmmoreno@ull.edu.es

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_1

find container vessels, which are aimed at moving containers among distant terminals along a predefined maritime route within container transportation networks. Also, land transportation means are brought together in a conventional maritime container terminal. In most of the cases, trucks and trains arrive at a maritime container terminal with the aim of picking-up and deliverying containers.

Broadly speaking, the freights are fundamentally moved along global supply chains through the use of standard-sized metal boxes: containers. The appearence of the container as international vehicle for the freights has constituted a true industrial revolution due to the fact that it encourages to reduce transportation costs by exploiting the benefits provided by the economies of scale, prevents cargo damages, and minimizes shipping costs. They are designed in accordance with predefined dimensions with the aim of enabling their exchange within multi-modal transportation networks in which at least vessels, trucks, and trains are brought together. It is worth mentioning that the standard capacity unit of measure for containers is the Twenty-foot Equivalent Unit, which measures about 6 m long.

The relevance of maritime container terminals has progressively increased from the time of the introduction of the containerization in the international trade trough the present time. Nowadays, maritime container terminals are facing with the increasing growth in the world merchandise trade and seaborne shipments. It must be pointed out that the seaborne shipments and the merchandise trade have increased in tandem over the last years. As indicative data, more than 10 billion tons of freights have been moved around the world by shipping business during 2015, according to the Review of Maritime Transport 2016 published by the United Nations Conference on Trade And Development (UNCTAD).[1] This constitutes a growth of 2.5% in the world seaborne trade. However, projections indicate that this volume of containers is going to be largely exceeded in the course of the coming years after the financial and economic crisis.

As previously indicated, in order to manage such a great volume of containers, large maritime container terminals are required. The maritime container terminals are multi-modal logistic interfaces dedicated to connect maritime and hinterland means of transportation. In this context, the layout of a conventional maritime container terminal consists of the following functional areas [40]:

- *Sea-side*. It is the part of the terminal in which the incoming container vessels are berthed. The main goal of this functional area is to carry out the transshipment of containers included into the corresponding stowage plan efficiently. For this purpose, a set of quay cranes available at the terminal is deployed. These cranes are aimed at unloading the containers whose destination is the terminal in which they are available. At the same time, they are also dedicated to load containers into the incoming vessels to be transported to other maritime container terminals. The book [71] provides a rigurous analysis of the main logistic operations in the sea-side.

[1] http://unctad.org.

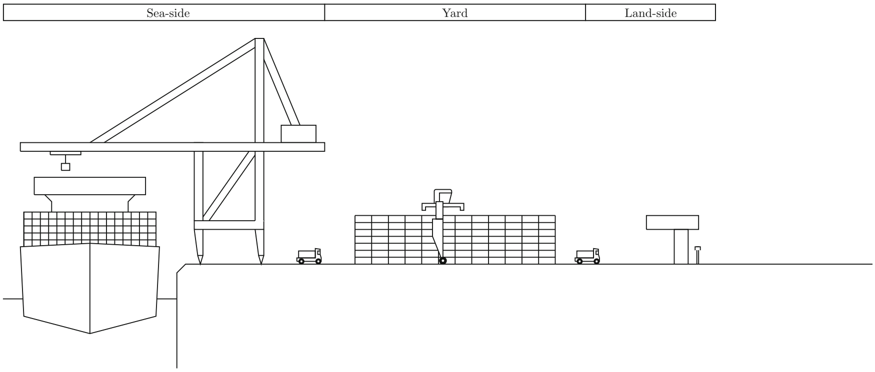


Fig. 1 Functional areas of a conventional maritime container terminal

- *Yard*. It is a large open-air surface aimed at storing containers temporarely until their later retrieval. The containers are distributed among homogeneous blocks, which are three-dimensional storages. Each block is divided into a set of parallel bays, which are composed of a set of contiguous stacks with a maximum stacking limit [9].
- *Land-side*. It is the interface dedicated to exchange freights between the maritime container terminal and the hinterland means of transportation. With this goal in mind, a set of external gates are available to monitor and control freights entering and leaving the infrastructure [10].

These functional areas of a maritime container terminal are illustrated from left to right in Fig. 1.

The technological advancement has given rise to the maritime container terminals have become complex infrastructures to manage due to the fact that its operating efficiency must be achieved through improved productivity and performance of analysts and transportation agents. However, the large and diverse number of logistic activities brought together is a major obstacle to the effective decision-making processes. In most of the cases, the logistic processes in this environment belong to the $\mathcal{NP}$-hard class of problems. Furthermore, practitioners must deal with the imprecision and uncertainty regarding the arrival and retrieval of containers, changing information, etc. The existence of uncertainty, imprecision, and partial truth in logistic processes must be explicitly taken into account as much as possible. Unfortunately, the narrow scope in practice of hard computing approaches to handle these issues in an appropriate form encourages the study and implementation of soft computing proposals.

The present chapter is aimed at reviewing the main logistic processes arisen in the context of maritime container terminals and its related fields. Specifically, we present and illustrate the most outstanding optimization problems studied by the scientific community over the last few decades and discuss the main findings associated with the existing research of soft computing methodologies. In this regard, our main objective is to provide a general overview of the most relevant solving approaches related to the logistic processes from the perspective of soft computing and identify some opportunities to go further into depth on knowledge.

The remainder of this chapter is organized as follows. Firstly, Sect. 2 describes the main logistic processes associated with the movement of freights at maritime container terminals. Afterwards, Sect. 3 overviews the major contributions made by the soft computing to maritime logistics. Finally, this chapter ends by a brief summary and indicates several promising lines for further research in maritime logistics and its related fields.

2 Logistic Processes

In spite of the fact that all the means of transportation can be present, maritime container terminals play a relevant role within global supply chains due to the fact that they are able to connect the maritime and hinterland means of transportation efficiently. Their primary goal is to enable the exchange of freights between producers and consumers with no repacking. The global supply chains in which the implicated maritime container terminals are inherently complex to manage due to the large amount of containers, technical characteristics of the loading and unloading equipment, existing staff, and legal aspects, among others.

From a general perspective, a global supply chain is composed of multiple flows of standardized containers in which the freights are packaged into. Before loaded with freights, the containers are manufactured by some metal factory from weathering steel, powder coated a color, and with the dimensions imposed by the requirements of the carriers. The empty containers are temporarily stored on a depot until the producers of the freights to transport request them. Once required, an empty container is transported by road to the producer of the freights, or some infrastructure dedicated to this goal, to move toward their consumers. The containers stuffed with the cargo to transport are moved to some maritime container terminal to be shipped overseas by some container vessel. In the simplest case, one truck moves each container to the terminal. However, multiple hinterland transportation means and container terminals can be involved in this process. The containers are stored on the yard of the terminal until the target vessels that are actually at the terminal are ready to load them onto. Similarly as in the hinterland case, the oversea transportation can involve multiple vessel journeys. Some minor container terminals, termed hubs, are usually used as intermediate storage infrastructures when moving containers over long distances. The target maritime container terminal is considered as the departure point of the hinterland transportation toward the customers of the freights. Lastly, the freights are stripped and the empty containers are stored again in a depot until another transportation is requested.

The efficient exchange of freights between heterogeneous means of transportation in a maritime container terminal depends on having a suitable integrated planning, coordination, execution, and control of container flows. In this scenario, the source and destination of each container define its type. Specifically, an import container is that unloaded from an incoming container vessel or delivered by a truck or train and must be stored on the yard of the terminal. Similarly, an export container is that stored on the yard of the terminal and must be included into the stowage plan of some incoming container vessel or loaded onto a truck or train. Lastly, a transshipment container is that unloaded from an incoming container vessel, stored on a block of yard, and transferred to another incoming container vessel to be moved toward its destination port.

The maritime container terminals are large engines for the economic development because they connect production sources and end consumers. In this regard, a maritime container terminal is responsible for serving efficiently the means of transportation brought together. However, the high complexity associated with the management of the container flows arisen in these infrastructures poses a continuing challenge that has to be recognised and addressed by practitioners. In particular, multitude of logistic planning problems known as $\mathcal{NP}$-hard are identified in a conventional maritime container terminal.

Several general classifications have been already proposed in the scientific literature to organize the logistic planning problems according to the functional area in which they take place, the means of transportation involved in them, or the planning horizon considered, among others. In this chapter, we use the classification of the logistic problems proposed in [43]. In this case, the logistic problems are split into four categories on the basis of the functional areas. These categories are termed *ship to shore*, *transfer*, *storage*, and *delivery*. The most highlighted optimization problems included into the previous categories are summarized in the following:

- *Ship to shore*. It involves those optimization problems associated with the loading and unloading operations of containers included into the stowage plan of the incoming vessels:
  - *Ship routing* [54]. It seeks to determine the best journeys of a fleet of container vessel to exchange containers among a given set of maritime container terminals.
  - *Stowage planning* [74]. Its objective is to identify the positions assigned to the containers to carry by a vessel while taking into account the features of its shipping route.
  - *Berth allocation* [7]. It seeks to determine the berthing position and berthing time of the incoming container vessels arrived to the maritime container terminal to minimize their service times.
  - *Quay crane assignment* [72]. This optimization problem is aimed at determining the subset of quay cranes to assign to the incoming container vessels over a given planning horizon to provide the best possible service.

   - *Quay crane scheduling* [70]. Its goal is to obtain a schedule of loading and unloading operations of containers to perform by the quay cranes assigned to a particular container vessel in such a way that its waiting time is minimized.

- *Storage*. It involves those optimization problems arisen from the movement and warehousing of containers on the yard and its related facilities to be ready for loading onto the incoming means of transportation:
   - *Yard crane scheduling* [36]. It is aimed at determining an efficient schedule of the storage, retrieval, and relocation of containers around the yard of the terminal.
   - *Container storage* [12]. It is a family of optimization problems aimed at maximizing the performance of the stacking cranes on the yard when storing and retriving containers.

- *Transfer and delivery*. It comprises those optimization problems arisen from the movement of containers between the functional areas of the terminal:
   - *Vehicle dispatching* [3]. It is aimed at managing the internal delivery vehicles of the terminal to maximize their performance.
   - *Gate operations planning* [17]. It seeks to optimize the access of the trucks and trains to the terminal to fulfill access and capacity restrictions.

In the remainder of this chapter, the focus is put on the application of soft computing methodologies in maritime logistics and its related fields.

3 Soft Computing in Maritime Logistics and Related Fields

Most of the authors who have addressed planning problems in maritime container terminals and its related fields over the last decades have assumed a very optimistic standpoint, in which uncertainty is completely ruled out. However, despite of the fact that the scientific literature considers the planning problems arisen in maritime container terminals as completely static in most of the cases, multitude of uncertainty sources are present in real-life scenarios. In particular, several general types of uncertainty sources can be distinguished. The former is mainly composed of that uncertainty derived from those elements in which an accident could happen in the infrastructure, information resources, communication systems, machinery, staff, or environment, among others. This is the case of a traffic accident in which a stacking crane is directly involved. Also, a second type of uncertainty source is composed of those elements that give rise to a change of the freight requests. Illustrative examples of this type of uncertainty are the potential wide-ranging fluctuations in the arrival or departure of container vessels, late retrieval of containers by external trucks, among others.

The existence of such a complex and dynamic environment within global supply chains forces terminal managers to be provided with efficient tools to manage it. The reason is found in that considering uncertainties in the scenario appropriately allows to mitigate the impact of potential disruptions, as well as the infeasibility of the working operations over a given planning horizon. In this regard, pro-active and reactive approaches can be considered. Obtaining robust solutions of the planning problems under analysis is the main goal of pro-active approaches, whereas reactive approaches are aimed at recovering the complete infrastructure when some unforeseen incident has already happened. A classic example of pro-active approaches is the insertion of slack time when scheduling container requests in such a way that the arrival of new requests can be easily assumed by the terminal with the lowest possible impact on the overall service quality. The major drawback associated with these approaches is the time that is not spent in effective work due to the existence of slack time in the schedules.

Since the introduction of the worldwide accepted definition of soft computing by Zadeh [101] in 1994, the use of soft computing methodologies has consolidated as a relevant branch of science to be applied in maritime logistics and its related fields. In this regard, soft computing has gained in importance due to the fact that it has the ability to tackle optimization problems classified as complex according to the complexity theory and in which partial truth, imprecision, and uncertainty appear and with the aim of achieving low cost solutions, tractability, and truth. For this reason, successful soft computing applications have attracted increased attention by the research and practitioner communities over the last decades. However, the large volume of research published so far encourages to make available classifications and reviews of the most highlighted works to draw the main contributions and identify promising lines for further research.

According to the previous discussion, in the remainder of this section we review the main applications of soft computing methodologies in maritime container terminals and its related fields. This review is organized on the basis of the main components of the soft computing, described in [93]. These components are briefly described in the following:

- *Probabilistic models*. It handles stochastic uncertainty. That is, the uncertainty derived from the potential occurrence of a particular event is quantified by a certain degree of probability. The most extended methods in this field are based on the Bayesian calculus, which allows to consider probability statements.
- *Fuzzy logic*. It is a mathematical tool that allows to capture the tolerance to errors and is aimed at dealing with approximate reasoning in which fuzzy truth-values are used as adapter elements applied to fuzzy statements.
- *Artificial neural networks*. They are systems integrated by multiple simple processing components that work in parallel with the goal of exhibiting some brain-like behavior.
- *Metaheuristic techniques*. They are computational techniques designed to provide approximate solutions of a given optimization problem and which largely fulfill the requirements of the decision maker.

3.1 Approximate Reasoning

3.1.1 Probabilistic Models

The probabilistic models result from the need of tackling jointly vagueness and probability associated with real-life problems. In the context of maritime container terminals and its related fields, its is usual to address planning problems in which it is required to reason in presence of uncertain information under partial knowledge. Terminal manager demand to count on intelligent tools to support in the identification of those unforeseen events that could happen and to assess their potential impact and consequences. The probabilistic models are essential ingredients in effective approaches aimed at minimizing their occurrence.

The resilience of a maritime container terminal is quantified through a Bayesian network in [45]. As discussed by the authors, Bayesian networks are able to draw relationships among different variables involved in the performance of the whole infrastructure. Also, the container operator efficiency is assessed from the perspective of soft computing in [99]. In this case, the authors present an empirical model that measures efficiency changes and which is estimated by means of a Bayesian approach supported by a Markov chain Monte Carlo simulation to make inference of the unknown parameters.

Another representative example of a probabilistic model in the scientific literature applied to maritime logistics is presented in [1]. This paper introduces a model to apply a failure mode and effects analysis when evaluating the performance of safety measures integrated into the operational system of a container terminal. Specifically, this model combines a fuzzy rule-based Bayesian network with evidential reasoning. Their goals are to describe input failure information to identify hazardous events and to aggregate these events while enabling dynamic risk-based decision support, respectively. Furthermore, the container throughput forecasting in maritime terminals has been a topic which has traditionally attracted the attention of the research community. In this regard, [98] discusses the applicability of three hybrid approaches based on least squares support vector regression model for this goal. The proposed approaches are compared to each other and to benchmark proposals. The computational experiments indicate clearly that seasonal decomposition of the series is an effective approach to obtain a good container throughput forecast.

Moreover, a lateness probability of the containers when moving around the different functional areas of a terminal is studied in [88]. In this case, the authors propose a Bayesian network designed to exploit the information recorded by the information processing systems of the terminal in the form of event logs. This network is built by considering the causal execution and co-occurrence between events to predict lateness probabilities.

3.1.2 Fuzzy Logic

Fuzzy logic is a mathematical formal multi-valued logic concept which is able to tackle imprecision, uncertainty, lack of information, and partial truth by imitating complex perception processes. This analytical tool has definitely played a leading role in maritime logistics over the last years when addressing planning problems for which hard computing methodologies have not proved successful. However, the implementation of approaches based upon fuzzy logic has been particularly remarkable when controlling driving-related tasks associated with technical machinery. Some examples are the guidance system and steering control of internal delivery vehicles, the real-time illumination stability system of stacking cranes, among others. Input data are expressed by means of linguistic variables in control applications, and taken together with if-then statements are used to formulate the conditional cases and eventually to produce representations of human knowledge. In addition, the optimization criteria of the planning problems arisen in maritime logistics are usually difficult to be defined accurately due to the fact that there is a certain inherent degree of imprecision in the way the preferences, constraints, and priorities of stakeholders are expressed.

One of the applications of fuzzy logic with the highest impact on maritime logistics is related to the control of quay and gantry cranes. The larger the cranes, the greater the needs of controlling them while satisfying strict specifications about payload position and swing angles, among others. A fuzzy logic-based controller is presented in [86] to avoid payload oscillations. Other examples of fuzzy logic in maritime logistics are related to the competitiveness of the terminals. The competitiveness of any maritime container terminal is highly influenced by its capability to attract shipping lines and retain those it is serving so far. In this regard, a large amount of factors are behind the choice of a port by a maritime operator that must be appropriately quantified. Some of these are the connectivity, efficiency, port charges, and range of port services, among others. The paper [100] presents a fuzzy evidential reasoning method to choice ports under uncertain environments from the perspective of shipping lines. The evidential reasoning is here used to values associated with the factors involved in the port alternatives. In fact, objective and subjective data are computed as fuzzy grades through linguistic terms based upon certain degrees of belief. These are later combined by using evidential reasoning to obtain the assessment of the alternatives. The computational experiments carried out in the paper under analysis indicate that the proposal is able to ease the exhaustive assessment of ports. The work [32] introduces fuzzy mathematical models aimed at determining the berthing time of incoming container vessels in a terminal and scheduling their transshipment operations. The arrival time of the containers and the processing time of the transhipment operations are considered as fuzzy. Another interesting application of fuzzy logic in maritime logistics is discussed in [39]. The authors present a probabilistic-fuzzy method that allows to determine quantitatively the probability of dangerous situation occurrence of a vessel manoeuvring in waterways by taking into account scenarios in which navigational safety is threatened. Furthermore, [86] presents a fuzzy logic-based robust feedback anti-sway control system which can be used either with or without a sensor of sway angle of a payload. As described by the authors, unlike other fuzzy approaches based on linguistic rule-based strategies and tuning of membership functions, the cited paper considers an interval analysis of closed-loop control system characteristic polynomial coefficients to solve the fuzzy interpolation control scheme design.

3.2 Functional Approximation and Optimization Methods

3.2.1 Artificial Neural Networks

Since the introduction of the seminal computational model inspired by the human brain in 1943 [69], the study of artificial neural networks has consolidated itself as an essential element in the field of soft computing. Artificial neural networks are known to be highly efficient to simulate the learning processes of human brains by miming the biological neurons in a nervous system. The group of interconnected artificial neurons in a neural network is in this case considered as a combination of simple processing elements that allows to provide advanced reasoning.

The application of artificial neural networks in the context of maritime container terminals has expanded dramatically in recent years due to their ability for reporting approximation to analytical functions, describing the behavior of carriers, containers, staff, among others, as well as predicting recurring phenomena derived from time series. This ability is especially interesting in maritime logistics because mathematical descriptions of the planning problems are not always available to be addressed and complex relations among variables are usually present. Instead, a large amount of data is usually reported by the stakeholders, from which neural networks can learn the underlying model and the interdependencies among parameters to support decision making processes.

Artificial neural networks have been successfully applied in maritime logistics and its related fields in manifold ways. Some of these are container demand forecasting, freight control, assessment of transportation parameters, maintenance of logistic infrastructures, among others. As an example, an artificial neural network model is proposed in [66] to determine the wave agitation in Spanish ports in order to provide a suitable anchorage for the incoming vessels. In particular, multilayer feed-forward back-propagation neural networks are considered due to the generalization capabilities to estimate wave heights. The Levenberg-Marquardt algorithm is here used to train the network, whereas a Bayesian regularisation is integrated to avoid over-fitting. The proposed model uses time series of deep-water wave buoy observations alone to obtain new knowledge and overcome the drawbacks associated with previous physical and numerical models used traditionally for this purpose. The computational experiments demonstrate that, unlike classic approaches, the proposal is more simple, does not require a large amount of data, and has a more efficient performance. Another application of artificial neural network to the reliability of coastal structures is described in [47]. In this case the authors propose a model based on an artificial neural network to estimate the armor damage sustained by a rubble-mound breakwater under wave action. A similar approach is presented in [50]. Specifically, a combination of artificial neural network and a Monte Carlo simulation is used to estimate damage of breakwater armor blocks. Other interesting examples of artificial neural network in wind direction forecasting [89], wave forecasting [25], long-wave prediction inside the port [67], or port tranquility [65], have been published in the scientific literature.

3.2.2 Metaheuristic Techniques

Since the term metaheuristic was coined by Fred Glover in 1986, a wide variety of definitions have been provided in the scientific literature. The paper [87] defines "a metaheuristic as a high-level problem-independent algorithmic framework that provides a set of guidelines or strategies to develop heuristic optimization algorithms. The term is also used to refer to a problem-specific implementation of a heuristic optimization algorithm according to the guidelines expressed in such a framework."

Most of the optimization problems that take place in maritime container terminals belong to the $\mathscr{NP}$-hard class of problems. Even in small-size scenarios of some of these optimization problems, efficient exact approaches do not exist. Therefore, metaheuristic algorithms provide high-quality solutions within short computational times. Taking into account the classification of these problems given in [43] and described in Sect. 2, in the following, the application of metaheuristics to some of the most highlighted optimization problems within the categories *Ship-to-shore* and *Storage* are briefly summarized. Genetic Algorithms (GA) [44], GRASP [33], Tabu Search (TS) [38], Variable Neighborhood Search (VNS) [73], Large Neighborhood Search (LNS) [85], Adaptive Large Neighborhood Search (ALNS) [83], and Simulated Annealing (SA) [53] are among the most effective algorithms to solve the considered optimization problems. However, due to the very large volume of publications, the following review is limited to indicate those publications appeared over the last years and considered as representative in the view of the authors of this chapter.

- *Ship-to-shore*.
  - *Ship routing and scheduling*.
    The goal of this class of optimization problems is to determine the best journeys of a fleet of container vessel to exchange containers among a given set of maritime container terminals. Depending on the operation mode, three kinds of ship routing problems can be distinguished: liner, industrial, and tramp shipping. Liners operate according to an agreed itinerary and schedule similar to a bus line. In industrial shipping, the cargo owner or shipper controls the ships. Industrial operators strive to minimize the costs of shipping their cargoes. Tramp fleets engage in contracts to transport specified (usually large) volumes of cargo between two ports within a period of time. They engage in contracts to make one or several trips, each trip having specified origin and destination ports and time windows for picking and delivering the cargo. Tramp is usually the

operation mode selected to transport liquid and dry commodities, or cargo involving a large number of units.

The fast growth of the containership fleet has resulted in a large number of research papers about liner network design and related topics published in the last decades. Recent reviews about ship routing problems can be found in works [19, 20], where literature contributions are classified. Most papers about ship routing and scheduling problems focus on the development of Mixed Integer Programming (MIP) models or heuristic/metaheuristic methods to solve them. Given the fact that the works by Christiansen et al. provide a comprehensive survey about ship routing and scheduling problems, this review is limited to some of the most recent literature works that present metaheuristic approaches for solving this kind of problems.

The work [6] proposes an Adaptive Large Neighborhood Search heuristic for a ship routing and scheduling problem with voyage separation requirements. [75] proposes a Genetic Algorithm with Local Search to solve a ship routing problem. [84] combines a Simulated Annealing with a Genetic Algorithm, and [54] uses a Tabu Search algorithm which allows infeasible solutions with respect to ship capacity and time. Other works in the literature introduce specific concepts in ship routing problems. [82] proposes a GRASP and discusses aspects related to data gathering and updating, which are particularly difficult in the context of ship routing. Lastly, [55] considers a cost function that depends on the wind speed and its direction, as well as on the wave height and its direction, and solves the problem using a Simulated Annealing algorithm.

– *Stowage planning*. The goal of this class of problems is to determine the position to be occupied by each container into a vessel taking into account the shipping route of that vessel. Notice that once a vessel arrives at a port, shifts are not desired. Shifts are the movements that correspond to the temporarily unloading and re-loading of containers in order to retrieve other containers that have to be unloaded at that port. In addition, any stowage plan has to lead to a seaworthy vessel, whose static stability is correct and all stress forces are within some limits as stated in [18]. [78] states that there are two main approaches to solve the ship stowage planning problem (SSPP): single-phase approaches, which tackle the SPP as a whole, and two-phases approaches, which consider a hierarchical decomposition of the problem - master planning, that assigns the containers to locations of the vessel, and slot planning, that determines the exact position of a container within a location.

  Since the publication of the paper [5] in 1993, in which exact and heuristic solutions were proposed for solving the SSPP in order to minimize the number of shiftings without considering stability constraints, many variants of the stowage planning problem using either a single-phase or a two-phase approach have been solved by means of metaheuristics. Genetic algorithms, GRASP, Tabu Search, and Simulated Annealing are among the most effective algorithms for the SSPP. The work [102] presents a Genetic Algorithm based on NSGA−III combined with a local search component to solve a multiobjective SSPP with the goals of optimizing the ship stability and the number of re-handles. [103] presents a Genetic Algorithm to tackle the SSPP for 40-feet outbound containers. [80] presents a GRASP algorithm to solve a generalization of the Slot Planning Problem, in which the explicit handling of rolled out containers and separations rules for dangerous cargo are introduced. [4] proposes the use of the hybrid method Pareto Clustering Search, which combines metaheuristics based on Simulated Annealing and local searches to solve the 3D Container ship Loading Plan Problem. [74] proposes a two-step heuristic for solving the SSPP, which are both based on the Tabu Search metaheuristic. [23] proposes a combination of metaheuristics, including Genetic Algorithm and Simulated Annealing, to solve the 3D container ship loading planning problem. [2] presents a three-step heuristic for the Master Bay Plan Problem, making use of tabu search to look for the global ship stability of the overall stowage plan. [49] proposes a Genetic Algorithm to solve the simultaneous stowage and load planning for a container ship with container rehandle in yard stacks with two objectives, ship stability and the minimum number of shifts. [26] also proposes a Genetic Algorithm for solving the SSPP with the goal of minimizing the number of container movements. The work [97] also presents a Tabu Search algorithm to solve the SSPP.

– *Berth allocation and Quay crane assignment*.
Given a berth layout and a set of vessels to be served, the aim of the Berth Allocation Problem (BAP) is to determine a berthing time and a berthing position for each vessel in order to optimize a given objective function. The scientific literature is replete with variants of this problem, which depend mainly on the berth layout (i.e., discrete, continuous, and hybrid), the arrival times of the vessels (i.e., static and dynamic) and the optimization level (i.e., strategic, tactical, and operational, which is the most extended in the literature). Taking into account the berth layout, a discrete quay is divided into several berths, in which a single vessel at a time can be served. In a continuous quay, the vessels can be assigned to any position as far as no space-time overlaps appear. In the hybrid case, the quay is also divided into berths, but the vessels can share them. Moreover, depending on the arrival times of the vessels, in the static case, all the vessels are in port before the planning horizon, while in the dynamic case, they can arrive at any time during the planning horizon. Finally, the optimization level can be either operational, when it covers decisions ranging from one up to several days, tactical, when the decisions cover operations ranging from one week up to several months, and strategic, when the decisions range from one up to some years. In addition, several objective functions have been considered in the literature. The main goal is to optimize the delays and waiting times of container vessels at the operational level. At this level, some of the goals are to optimize the transshipment flows among vessels, cycling visiting of the vessels, fulfillment of contracts among shipping companies and terminal managers, route design, etc. At the strategic level, the problem seeks to establish specific and dedicated berths, strategic cooperation agreements between terminal and shipping companies, etc. Note that the Quay Crane Assignment Problem, whose aim is to assign a set of quay cranes to a vessel to perform the loading and unloading operations, is usually solved together with the Berth

Allocation Problem. Therefore, it is integrated in the variants that are related in the literature.

If attention is focused on journal papers, it can be checked that, since the publication of the paper [48], less than three papers per year related to variants of berth allocation were published until 2006. This number rapidly increased to reach more than fifteen papers published in 2016. Such as it is reported in the survey [8], a large number of meta-heuristics have been effective to solve this kind of optimization problem. Genetic Algorithms, Tabu Search, Simulated Annealing, and GRASP are among the most widely used techniques.

Given the fact that the work [8] provides a comprehensive review about BAP, the following is limited to cite some of the papers by the authors of this book chapter related to different variants of BAP. The work [56] proposes a cooperative search to solve the discrete dynamic BAP. In the work [59], the authors consider an aditional constraint that appear in real situations, water depth and tidal constraints. It is proposed a POPMUSIC approach (Partial Optimization Metaheuristic Under Special Intensification Conditions) that includes the resolution of an appropriate mathematical programming formulation as an embedded procedure. The paper [57] proposes a Biased Random Key Genetic Algorithm for solving the tactical Berth Allocation Problem. Finally, the work [58] presents a hybrid Tabu Search—Path Relinking to solve the dynamic BAP.

– *Quay crane scheduling*.

  It is an optimization problem arisen to deal with the transshipment operations associated with each incoming container vessel arrived at a maritime container terminal. In particular, the stowage plan of a vessel indicates the individual containers to be loaded and unloaded onto/from it after its berthing. These containers must be handled by a subset of the quay cranes available at the terminal. Because crane operations are highly expensive, terminal practitioners must determine a suitable schedule in order to deliver high quality service while reduce operative costs.

  The Quay Crane Scheduling Problem, in short QCSP, has attracted a great deal of interest of the soft computing community due to the fact that it introduces a set of novel and challenging constraints in comparison with other scheduling problems tackled to date. This is the case of, for example, the well-known Job Shop Scheduling Problem. On one hand, the movements of the quay cranes used to perform the transshipment operations are physically restricted because these are mounted on a system of rails. This means that the cranes can only move horizontally along the berthing line. Also, the quay cranes cannot cross to each other, which gives rise to their initial relative order is kept over the planning horizon. Lastly, the cranes have to keep a safety distance between them in order to prevent potential collisions.

  Multitude of proposals belonging to the field of soft computing have appeared in the scientific literature to address the QCSP. These approaches can be classified on the basis of different criteria. Some of these are the level of aggregation of the containers into the stowage plan to handle, the technichal characteristics of the machinery and their crane drivers, level of potential interferences, and performance measure. Firstly, the containers into the stowage plan of a vessel are usually arranged according to their destination port or physical characteristics, among others. This way, containers with similar characteristics can be easily loaded/unloaded in a row at a given bay of the vessel. The level of aggregation indicates the volume of containers to handle. At the lowest level of aggregation, individual containers are handled by the quay cranes, whereas all the containers in a bay are handled at the highest level. Representative metaheuristics for this variants can be found in the works [41, 90], respectively. Furthermore, one of the most complex constraints of the QCSP is associated with the operative chracteristics of the quay cranes. In this context, some metaheuristics have been designed to address the impact of the temporal availability of the cranes on the overall performance of the transshipment operations. This is the case of the work [62]. However, other authors have also proposed efficient metaheuristics in which the movement of the quay cranes is considered as non-negligible, as stated in the work [68]. In addition, the potential interference derived from the presence of a rail system has been an attractive focus of research. A safety distance is considered by some metaheuristics to avoid collisions between quay cranes. For example, this is considered in [21, 27]. Finally, a wide range of performance measures have been already addressed in the scientific literature so far. In particular, some high-efficient metaheuristics have been designed to minimize the completion time of the operations [64], the finishing times of the quay cranes [22], the movement of the quay cranes [96], and maximize the crane utilization rate [95], among others. It is worth pointing out that a comprehensive review of the proposals aimed at solving the QCSP can be found in the survey [8].

- *Storage*.
  - *Yard crane scheduling*
    Given the fact that the shipping lines are the customers of a container terminal and their payment depends on the time spent at the terminal, the main goal of the berth allocation problem usually considers minimizing the berthing time of the vessels. In order to minimize this time, the loading and unloading operations have to be efficiently performed taking into account the available resources and their particular efficiencies. Quay cranes, internal vehicles, and yard cranes play then a crucial role. Although the speed of quay cranes has improved substantially over the last years, their performance depends on the performance of yard cranes, whose speed is approximately one-third of a modern quay crane. Therefore, in order to fully utilize quay cranes, containers to be loaded or unloaded by them are distributed over several blocks. Moreover, while retrieving and stacking containers of ships, each yard crane has also to serve the landside. Then, the goal of the Yard Crane Scheduling Problem is to determine an efficient schedule to carry out a set of container storage, retrieval and relocation requests around the yard of the terminal as indicated in [36].
    In the following, a reduced number of the recent references related to the Yard Crane Scheduling Problem. [42] proposes the combination of a Genetic Algo-

rithm with Particle Swarm Optimization to solve the Yard Crane Scheduling Problem with the aim of improving the efficiency of the terminal and minimizing energy-consumption. The authors of [94] solve the problem with two cooperating automated stacking cranes in a single block using a mathematical model and a Simulated Annealing based heuristic. [63] tackles the Yard Crane Scheduling Problem with inter-crane interference, fixed yard crane separation distances, and simultaneous container storage/retrievals by means of heuristics. The problem of scheduling multiple yard cranes by means of a dynamic programming-based heuristic and of an algorithm to find lower bounds is solved in [77]. [76] first carries out a theoretical investigation of the problem and then proposes a branch-and-bound based enumerative method and heuristics for solving it. Most of the works that tackle the problem at hand propose mathematical formulations to either solve it to optimality when it is possible or to obtain lower bounds or heuristics/metaheuristics to provide high quality solutions in reasonable computational times.

– *Container storage*
  Container storage is a three-level problem that arises at maritime container terminals, which involves defining the yard layout and selecting the handling machinery to use at a strategic level, determining the storage capacity of the yard and the handling machinery deployment at a tactical level, and moving the containers on the yard in a short-term at an operational level. At this last level, during a certain planning horizon, containers arrive and leave the yard at certain arrival and retrieval times, respectively. The objective of the yard cranes is to perform feasible movements to store and retrieve the containers on the basis of the intrinsic Last In First Out (LIFO) structure of the stacks. Therefore, a storage movement involves the placement of the next incoming container at the top of a stack with at least one empty slot, a retrieval movement involves taking out the next container to retrieve from its bay whenever it is currently placed at the top of a stack, and a relocation movement involves the movement of a container from the top of its assigned stack to the top of another one with at least one empty slot.
  In the state-of-art of container storage, the incoming and outgoing containers in a bay give rise to the definition of the following closely-related $\mathcal{NP}$-hard optimization problems:
  · *Stacking Problem* [28]. It is aimed at determining the shortest sequence of movements to be performed by the crane in order to store and retrieve the containers in/from the bay. The works [24, 51] estimate the number of relocation movements that are required to retrieve a random container from its current location. Also, [52] proposes a decision rule to determine the storage locations of relocated containers and to determine the containers to be retrieved among multiple containers with similar retrieval times. [79] derives formulas to estimate the number of relocation movements to retrieve a container from a stack using different stacking methods. [81] proposes several semi-greedy construction heuristics that are used in conjunction with a discrete-event sim-

ulation model to build feasible solutions for the stacking problem. Note that these works propose heuristics for solving the problem at hand.

- *Container Relocation Problem* [31]. It seeks to determine the shortest sequence of relocation movements to retrieve a subset of containers. It is assumed that all the containers are already stored in the bay and no new incoming containers arrive. [30] propose a heuristic algorithm aimed at solving the unrestricted blocks relocation problem considering the minimization of the number of relocation operations. [13] presents a complete study on the blocks relocation problem and propose an effective heuristic algorithm. [34] provides a general classification concerning all the feasible container relocation movements that can be performed on a given incumbent bay configuration. A tree search procedure is developed and a lower bound of the minimum number of relocation movements required to retrieve all the containers from the bay is considered to prune some branches of the tree. [16] addresses an extension of this problem in a storage area of the container yard where incoming and outgoing containers are arranged by a straddle carrier. They propose several constructive algorithms and three nature inspired metaheuristics are studied for improving the initial solutions reported by the heuristics. [15] presents a recursive formulation and a dynamic programming algorithm for the restricted blocks relocation problem and a corridor method. [61] presents a heuristic method composed of three general stages executed one after the other to retrieve the containers from a bay and move them toward a vessel. [11] proposes a smart binary encoding for the problem that allows to develop optimization methods without having in-depth knowledge concerning the current problem features. They design an algorithm based on the pilot metaheuristic, in which simple heuristics are included in order to compute the suitability of neighbor stacking configurations. [52] proposes two optimization methods to be applied when a container pickup sequence is given and use a heuristic rule to determine the number of expected future container relocation movements.
- *Pre-Marshalling Problem* [29]. Its objective is to find the shortest sequence of movements to arrange the containers within a given bay, such that any container is placed above other container with earlier retrieval time in the same stack. In this case, neither incoming containers nor outgoing containers are considered. [91] solves the pre-marshaling problem to optimality using A* and IDA*. [46] designes a biased random-key genetic algorithm. [37] proposes a variable chromosome length genetic algorithm. [92] proposes two metaheuristics, a Pilot method and a Max-Min Ant System, to solve the two-dimensional pre-marshalling problem. [29] proposes a heuristic algorithm and an instances generator for the pre-marshalling problem. [14] proposes a metaheuristic approach based on the paradigm of the Corridor Method. [60] designes an algorithm composed of a neighborhood search process, an integer programming model, and three minor subroutines.

4 Summary and Further Research

Container terminals are huge complex infrastructures located within the boundaries of maritime ports. The complexity arises from the large number of heterogeneous processes and multitude of stakeholders with conflicting goals that coexist in this context. For this reason, terminal managers demand nowadays to count with efficient operative strategies due to the fact that these allow to achieve the established performance and service objectives.

Up to now, most of the computational proposals aimed at improving the overall performance of maritime container terminals assume data is not influenced by uncertainty, inconsistency, nor noise. However, for example, multitude of uncertainty sources appear in a realistic environment. For this reason, intelligent approaches belonging to the field of soft computing have gained much popularity over the last decades when tackling problems in maritime logistics and its related fields. In this regard, the chapter at hand provides a general review of the most representative contributions of soft computing in this context. In particular, a brief overview of the main applications of soft computing methodologies is here presented. This applications are organized on the basis of the main components of the soft computing: probabilistic models, fuzzy logic, artificial neural networks, and metaheuristic techniques.

Finally, it is worth mentioning that despite the efforts done so far, there is still a large number of open promising lines for further research. One of these open lines involves the combination of both optimization techniques with online learning. This type of approaches are very useful to predict optimized process parameters associated with the movement of containers around the terminal. Also, hybrid proposals have demonstrated to be very effective when tackling complex optimization problems in which analytical techniques cannot be applied. Some of these proposals are genetic fuzzy systems and neural networks combined with evolutionary techniques, among others. The former are usually designed to obtain a suitable accuracy tradeoff in optimization cycles, whereas the last are able to solve multi-objective approximation.

Acknowledgements This work has been partially funded by the Spanish Ministry of Economy and Competitiveness (project TIN2015-70226-R) with FEDER funds.

References

1. Alyami, H., Yang, Z., Riahi, R., Bonsall, S., Wang, J.: Advanced uncertainty modelling for container port risk analysis. Accid. Anal. Prev. (2016). doi:10.1016/j.aap.2016.08.007
2. Ambrosino, D., Anghinolfi, D., Paolucci, M., Sciomachen, A.: A new three-step heuristic for the master bay plan problem. Marit. Econ. Logist. **11**(1), 98–120 (2009)
3. Angeloudis, P., Bell, M.G.H.: An uncertainty-aware AGV assignment algorithm for automated container terminals. Transp. Res. Part E: Logist. Transp. Rev. **46**(3), 354–366 (2010)

4. Araújo, E.J., Chaves, A.A., de Salles, L.L.: Neto, and A.T. de Azevedo. Pareto clustering search applied for 3D container ship loading plan problem. Expert Syst. Appl. **44**, 50–57 (2016)
5. Avriel, M., Penn, M.: Exact and approximate solutions of the container ship stowage problem. Comput. Ind. Eng. **25**(1–4), 271–274 (1993)
6. Bakkehaug, R., Rakke, J.G., Fagerholt, K., Laporte, G.: An adaptive large neighborhood search heuristic for fleet deployment problems with voyage separation requirements. Transp. Res. Part C: Emerg. Technol. **70**, 129–141 (2016)
7. Bierwirth, C., Meisel, F.: A survey of berth allocation and quay crane scheduling problems in container terminals. Eur. J. Oper. Res. **202**(3), 615–627 (2010)
8. Bierwirth, C., Meisel, F.: A follow-up survey of berth allocation and quay crane scheduling problems in container terminals. Eur. J. Oper. Res. **244**(3), 675–689 (2015)
9. Carlo, H.J., Vis, I.F.A., Roodbergen, K.J.: Storage yard operations in container terminals: Literature overview, trends, and research directions. Eur. J. Oper. Res. **235**(2), 412–430 (2014) (Maritime Logistics)
10. Carlo, H.J., Vis, I.F.A., Roodbergen, K.J.: Transport operations in container terminals: Literature overview, trends, research directions and classification scheme. Eur. J. Oper. Res. **236**(1), 1–13 (2014)
11. Caserta, M., Schwarze, S., Voβ, S.: A new binary description of the blocks relocation problem and benefits in a look ahead heuristic. In: Cotta, C., Cowling, P. (eds.) Evolutionary Computation in Combinatorial Optimization. Lecture Notes in Computer Science, vol. 5482, pp. 37–48. Springer, Berlin (2009)
12. Caserta, M., Schwarze, S., Voβ, S.: Container rehandling at maritime container terminals. In: Böse, J.W., Sharda, R., Voβ, S. (eds.) Handbook of Terminal Planning. Volume 49 of Operations Research/Computer Science Interfaces Series, pp. 247–269. Springer, New York (2011)
13. Caserta, M., Schwarze, S.: Voβ, S.: A mathematical formulation and complexity considerations for the blocks relocation problem. Eur. J. Oper. Res. **219**(1), 96–104 (2012)
14. Caserta, M., Voβ, S.: A corridor method-based algorithm for the pre-marshalling problem. In: Giacobini, M., Brabazon, A., Cagnoni, S., Di Caro, G.A., Ekárt, A., Esparcia-Alcázar, A.I., Farooq, M., Fink, A., Machado, P. (eds.) Applications of Evolutionary Computing. Volume 5484 of Lecture Notes in Computer Science, pp. 788–797. Springer, Berlin (2009)
15. Caserta, M.: Voβ, S., Sniedovich, M.: Applying the corridor method to a blocks relocation problem. OR Spectr. **33**(4), 915–929 (2011)
16. Casey, B., Kozan, E.: Optimising container storage processes at multimodal terminals. J. Oper. Res. Soc. **63**, 1126–1142 (2012)
17. Chen, G., Govindan, K., Yang, Z.: Managing truck arrivals with time windows to alleviate gate congestion at container terminals. Int. J. Prod. Econ. **141**(1), 179–188 (2013)
18. Christensen, J., Pacino, D.: A matheuristic for the cargo mix problem with block stowage. Transp. Res. Part E: Logist. Transp. Rev. **97**, 151–171 (2017)
19. Christiansen, M., Fagerholt, K., Nygreen, B., Ronen, D.: Ship routing and scheduling in the new millennium. Eur. J. Oper. Res. **228**(3), 467–483 (2013)
20. Christiansen, M., Fagerholt, K., Ronen, D.: Ship routing and scheduling: status and perspectives. Transp. Sci. **38**(1), 1–18 (2004)
21. Chung, S.H., Chan, F.T.S.: A workload balancing genetic algorithm for the quay crane scheduling problem. Int. J. Prod. Res. **51**(16), 4820–4834 (2013)
22. Chung, S.H., Choy, K.L.: A modified genetic algorithm for quay crane scheduling operations. Expert Syst. Appl. **39**(4), 4213–4221 (2012)
23. de Azevedo, A.T., Ribeiro, C.M., de Sena, G.J., Chaves, A.A., Neto, L.L.S., Moretti, A.C.: Solving the 3d container ship loading planning problem by representation by rules and metaheuristics. Int. J. Data Anal. Tech. Strateg. **6**(3), 228–260 (2014)
24. de Castillo, B., Daganzo, C.F.: Handling strategies for import containers at marine terminals. Transp. Res. Part B: Methodol. **27**(2), 151–166 (1993)

25. Deo, M.C., Jha, A., Chaphekar, A.S., Ravikant, K.: Neural networks for wave forecasting. Ocean Eng. **28**(7), 889–898 (2001)
26. Dubrovsky, O., Levitin, G., Penn, M.: A genetic algorithm with a compact solution encoding for the container ship stowage problem. J. Heuristics **8**(6), 585–599 (2002)
27. Expósito-Izquierdo, C., González-Velarde, J.L., Melián-Batista, B., Moreno-Vega, J.M.: Hybrid estimation of distribution algorithm for the quay crane scheduling problem. Appl. Soft Comput. **13**(10), 4063–4076 (2013)
28. Expósito-Izquierdo, C., Lalla-Ruiz, E., de Armas, J., Melián-Batista, B., Moreno-Vega, J.M.: A heuristic algorithm based on an improvement strategy to exploit idle time periods for the stacking problem. Comput. Ind. Eng. **87**, 410–424 (2015)
29. Expósito-Izquierdo, C., Melián-Batista, B., Moreno-Vega, J.M.: Pre-marshalling problem: heuristic solution method and instances generator. Expert Syst. Appl. **39**(9), 8337–8349 (2012)
30. Expósito-Izquierdo, C., Melián-Batista, B., Moreno-Vega, J.M.: A domain-specific knowledge-based heuristic for the blocks relocation problem. Adv. Eng. Inf. **28**(4), 327–343 (2014)
31. Expósito-Izquierdo, C., Melián-Batista, B., Moreno-Vega, J.M.: An exact approach for the blocks relocation problem. Expert Syst. Appl. **42**(1718), 6408–6422 (2015)
32. Expósito-Izquiero, C., Lalla-Ruiz, E., Lamata, T., Melián-Batista, B., Moreno-Vega, J.M.: Fuzzy optimization models for seaside port logistics: berthing and quay crane scheduling. In: Computational Intelligence, pp. 323–343. Springer (2016)
33. Feo, T.A., Resende, M.G.C.: A probabilistic heuristic for a computationally difficult set covering problem. Oper. Res. Lett. **8**(2), 67–71 (1989)
34. Forster, F., Bortfeldt, A.: A tree search procedure for the container relocation problem. Comput. Oper. Res. **39**(2), 299–309 (2012)
35. Fransoo, J.C., Lee, C.Y.: The critical role of ocean container transport in global supply chain performance. Prod. Oper. Manag. **22**(2), 253–268 (2013)
36. Gharehgozli, A.H., Yu, Y., de Koster, R., Udding, J.T.: An exact method for scheduling a yard crane. Eur. J. Oper. Res. **235**(2), 431–447 (2014) (Maritime Logistics)
37. Gheith, M., Eltawil, A.B., Harraz, N.A.: Solving the container pre-marshalling problem using variable length genetic algorithms. Eng. Optim. **48**(4), 687–705 (2016)
38. Glover, F.: Future paths for integer programming and links to artificial intelligence. Comput. Oper. Res. **13**(5), 533–549 (1986)
39. Gucma, L., Pietrzykowski, Z.: Ship manoeuvring in restricted areas: an attempt to quantify dangerous situations using a probabilistic-fuzzy method. J. Navig. **59**(02), 251–262 (2006)
40. Gunter, H.O., Kim, K.H.: Container terminals and terminal operations. OR Spectr. **28**, 437–445 (2006)
41. Hakam, M.H., Solvang, W.D., Hammervoll, T.: A genetic algorithm approach for quay crane scheduling with non-interference constraints at narvik container terminal. Int. J. Logist. Res. Appl. **15**(4), 269–281 (2012)
42. He, J., Huang, Y., Yan, W.: Yard crane scheduling in a container terminal for the trade-off between efficiency and energy consumption. Adv. Eng. Inf. **29**(1), 59–75 (2015)
43. Henesey, L.: Overview of transshipment operations and simulation. In: MedTrade Conference, Malta, pp. 6–7 (2006)
44. Holland, J.H.: Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. University of Michigan Press (1975)
45. Hosseini, S., Barker, K.: Modeling infrastructure resilience using bayesian networks: a case study of inland waterway ports. Comput. Ind. Eng. **93**, 252–266 (2016)
46. Hottung, A., Tierney, K.: A biased random-key genetic algorithm for the container premarshalling problem. Comput. Oper. Res. **75**, 83–102 (2016)
47. Iglesias, G., Rabuñal, J., Losada, M.A., Pachón, H., Castro, A., Carballo, R.: A virtual laboratory for stability tests of rubble-mound breakwaters. Ocean Eng. **35**(1112), 1113–1120 (2008)

48. Imai, A., Nagaiwa, K., Tat, C.W.: Efficient planning of berth allocation for container terminals in Asia. J. Adv. Transp. **31**(1), 75–94 (1997)
49. Imai, A., Sasaki, K., Nishimura, E., Papadimitriou, S.: Multi-objective simultaneous stowage and load planning for a container ship with container rehandle in yard stacks. Eur. J. Oper. Res. **171**(2), 373–389 (2006)
50. Kim, D.H., Kim, Y.J., Hur, D.S.: Artificial neural network based breakwater damage estimation considering tidal level variation. Ocean Eng. **87**, 185–190 (2014)
51. Kim, K.H.: Evaluation of the number of rehandles in container yards. Comput. Ind. Eng. **32**(4), 701–711 (1997)
52. Kim, K.H., Hong, G.P.: A heuristic rule for relocating blocks. Comput. Oper. Res. **33**(4), 940–954 (2006)
53. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**, 671–680 (1983)
54. Korsvik, J.E., Fagerholt, K.: A tabu search heuristic for ship routing and scheduling with flexible cargo quantities. J. Heuristics **16**(2), 117–137 (2010)
55. Kosmas, O.T., Vlachos, D.S.: Simulated annealing for optimal ship routing. Comput. Oper. Res. **39**(3), 576–581 (2012)
56. Lalla-Ruiz, E., Expósito-Izquierdo, C., Melián-Batista, B., Moreno-Vega, J.M.: A set-partitioning-based model for the berth allocation problem under time-dependent limitations. Eur. J. Oper. Res. **250**(3), 1001–1012 (2016)
57. Lalla-Ruiz, E., González-Velarde, J.L., Melián-Batista, B., Moreno-Vega, J.M.: Biased random key genetic algorithm for the tactical berth allocation problem. Appl. Soft Comput. **22**, 60–76 (2014)
58. Lalla-Ruiz, E., Melián-Batista, B., Moreno-Vega, J.M.: Artificial intelligence hybrid heuristic based on tabu search for the dynamic berth allocation problem. Eng. Appl. Artif. Intell. **25**(6), 1132–1141 (2012)
59. Lalla-Ruiz, E., Voß, S., Expósito-Izquierdo, C., Melián-Batista, B., Moreno-Vega, J.M.: A popmusic-based approach for the berth allocation problem under time-dependent limitations. Ann. Oper. Res. 1–27 (2015)
60. Lee, Y., Chao, S.L.: A neighborhood search heuristic for pre-marshalling export containers. Eur. J. Oper. Res. **196**(2), 468–475 (2009)
61. Lee, Y., Lee, Y.J.: A heuristic for retrieving containers from a yard. Comput. Oper. Res. **37**(6), 1139–1147 (2010)
62. Legato, P., Trunfio, R., Meisel, F.: Modeling and solving rich quay crane scheduling problems. Comput. Oper. Res. **39**(9), 2063–2078 (2012)
63. Li, W., Wu, Y., Petering, M.E.H., Goh, M., de Souza, R.: Discrete time model and algorithms for container yard crane scheduling. Eur. J. Oper. Res. **198**(1), 165–172 (2009)
64. Liu, M., Zheng, F., Li, J.: Scheduling small number of quay cranes with non-interference constraint. Optim. Lett. **9**(2), 403–412 (2015)
65. Londhe, S.N., Deo, M.C.: Wave tranquility studies using neural networks. Marine Struct. **16**(6), 419–436 (2003)
66. López, I., López, M., Iglesias, G.: Artificial neural networks applied to port operability assessment. Ocean Eng. **109**, 298–308 (2015)
67. López, M., Iglesias, G.: Artificial intelligence for estimating infragravity energy in a harbour. Ocean Eng. **57**, 56–63 (2013)
68. Lu, Z., Han, X., Xi, L., Erera, A.L.: A heuristic for the quay crane scheduling problem based on contiguous bay crane operations. Comput. Oper. Res. **39**(12), 2915–2928 (2012)
69. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**(4), 115–133 (1943)
70. Meisel, F.: Operational planning problems. Seaside Operations Planning in Container Terminals. Contributions to Management Science, pp. 17–30. Physica-Verlag, HD (2009)
71. Meisel, F.: Seaside Operations Planning in Container Terminals. Springer (2009)
72. Meisel, F., Bierwirth, C.: A framework for integrated berth allocation and crane operations planning in seaport container terminals. Transp. Sci. **47**(2), 131–147 (2013)

73. Mladenović, N., Hansen, P.: Variable neighborhood search. Comput. Oper. Res. **24**(11), 1097–1100 (1997)
74. Monaco, M.F., Sammarra, M., Sorrentino, G.: The terminal-oriented ship stowage planning problem. Eur. J. Oper. Res. **239**(1), 256–265 (2014)
75. Moon, I.K., Qiu, Z.B., Wang, J.H.: A combined tramp ship routing, fleet deployment, and network design problem. Marit. Policy Manag. **42**(1), 68–91 (2015)
76. Narasimhan, A., Palekar, U.S.: Analysis and algorithms for the transtainer routing problem in container port operations. Transp. Sci. **36**(1), 63–78 (2002)
77. Ng, W.C.: Crane scheduling in container yards with inter-crane interference. European Journal of Operational Research **164**(1), 64–78 (2005)
78. Pacino, D., Delgado, A., Jensen, R.M., Bebbington, T.: Fast generation of near-optimal plans for eco-efficient stowage of large container vessels. In: International Conference on Computational Logistics, pp. 286–301. Springer (2011)
79. Park, T.K., Kim, K.H.: Comparing handling and space costs for various types of stacking methods. Comput. Ind. Eng. **58**(3), 501–508 (2010)
80. Parreño, F., Pacino, D., Álvarez-Valdés, R.: A grasp algorithm for the container stowage slot planning problem. Transp. Res. Part E: Logist. Transp. Rev. **94**, 141–157 (2016) (cited By 0)
81. Rei, R.J., Pedroso, J.P.: Heuristic search for the stacking problem. Int. Trans. Oper. Res. **19**(3), 379–395 (2012)
82. Romero, G., Durán, G., Marenco, J., Weintraub, A.: An approach for efficient ship routing. Int. Trans. Oper. Res. **20**(6), 767–794 (2013)
83. Ropke, S., Pisinger, D.: An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. Transp. Sci. **40**(4), 455–472 (2006)
84. Santhanakrishnan, S., Narendran, T.T., Ganesh, K., Anbuudayasankar, S.P.: Comparison of meta-heuristics for container ship routing problem. Int. J. Serv. Oper. Manag. **12**(3), 348–367 (2012)
85. Shaw, P.: Using constraint programming and local search methods to solve vehicle routing problems. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **1520**, 417–431 (1998)
86. Smoczek, J.: Fuzzy crane control with sensorless payload deflection feedback for vibration reduction. Mech. Syst. Signal Process. **46**(1), 70–81 (2014)
87. Sörensen, K., Glover, F.: Metaheuristics. In: Gass, S., Fu, M. (eds.) Encyclopedia of Operations Research and Management Science, 3rd edn., pp. 960–970. Springer Science & Business Media (2013)
88. Sutrisnowati, R.A., Bae, H., Song, M.: Bayesian network construction from event log for lateness analysis in port logistics. Comput. Ind. Eng. **89**, 53–66 (2015) (Maritime logistics and transportation intelligence)
89. Tagliaferri, F., Viola, I.M., Flay, R.G.J.: Wind direction forecasting with artificial neural networks and support vector machines. Ocean Engineering **97**, 65–73 (2015)
90. Tang, L., Zhao, J., Liu, J.: Modeling and solution of the joint quay crane and truck scheduling problem. Eur. J. Oper. Res. **236**(3), 978–990 (2014)
91. Tierney, K., Pacino, D., Voβ, S.: Solving the pre-marshalling problem to optimality with a* and ida*. Flex. Serv. Manuf. J. 1–37 (2016)
92. Tus, A., Rendl, A., Raidl, G.R.: Metaheuristics for the two-dimensional container premarshalling problem. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **8994**, 186–201 (2015)
93. Verdegay, J.L., Yager, R.R., Bonissone, P.P.: On heuristics as a fundamental constituent of soft computing. Fuzzy Sets Syst. **159**(7), 846–855 (2008)
94. Vis, I.F.A., Carlo, H.J.: Sequencing two cooperating automated stacking cranes in a container terminal. Transp. Sci. **44**(2), 169–182 (2010)
95. Vis, I.F.A., van Anholt, R.G.: Performance analysis of berth configurations at container terminals. OR Spectr. **32**(3), 453–476 (2010)
96. Wang, S., Zheng, J., Zheng, K., Guo, J., Liu, X.: Multi resource scheduling problem based on an improved discrete particle swarm optimization. Phys. Proc. **25**, 576–582 (2012)

97. Wilson, I.D., Roach, P.A.: Principles of combinatorial optimization applied to container-ship stowage planning. J. Heuristics **5**(4), 403–418 (1999)
98. Xie, G., Wang, S., Zhao, Y., Lai, K.K.: Hybrid approaches based on LSSVR model for container throughput forecasting: a comparative study. Appl. Soft Comput. **13**(5), 2232–2241 (2013)
99. Yan, J., Sun, X., Liu, J.J.: Assessing container operator efficiency with heterogeneous and time-varying production frontiers. Transp. Res. Part B: Methodol. **43**(1), 172–185 (2009)
100. Yeo, G.T., Ng, A.K.Y., Lee, P.T.W., Yang, Z.: Modelling port choice in an uncertain environment. Marit. Policy Manag. **41**(3), 251–267 (2014)
101. Zadeh, L.A.: Fuzzy logic (abstract): Issues, contentions and perspectives. In: Proceedings of the 22nd Annual ACM Computer Science Conference on Scaling Up: Meeting the Challenge of Complexity in Real-world Computing Applications: Meeting the Challenge of Complexity in Real-world Computing Applications, CSC '94, New York, NY, USA, p. 407. ACM (1994)
102. Zhang, Z., Lee, C.Y.: Multiobjective approaches for the ship stowage planning problem considering ship stability and container rehandles. IEEE Trans. Syst. Man Cybern.: Syst. **46**(10), 1374–1389 (2016)
103. Zhao, N., Shen, Y., Xia, M., Jin, J.: A novel strategy for stowage planning of 40 feet containers in container terminals. J. Marine Sci. Technol. (Taiwan) **24**(1), 61–74 (2016)

Intelligent Data Analysis, Soft Computing and Imperfect Data

Jose M. Cadenas and M. Carmen Garrido

Abstract In different real problems the available information is not as precise or as accurate as we would like. Due to possible imperfection in the data (understanding that these contain data where not all the attributes are precisely known, such as missing, imprecise, uncertain, ambiguous, etc. values), tools provided by Soft Computing are quite adequate, and the hybridization of these tools with the Intelligent Data Analysis is a field that is gaining more importance. In this paper, first we present a brief overview of the different stages of Intelligent Data Analysis, focusing on two core stages: data preprocessing and data mining. Second, we perform an analysis of different hybridization approaches of the Intelligent Data Analysis with the Soft Computing for these two stages. The analysis is performed from two levels: If elements of Soft Computing are incorporated in the design of the method/model, or if they are also incorporated to be able to deal with imperfect information. Finally, in a third section, we present in more detail several methods which allow the use of imperfect data both for their learning phase and for the prediction.

1 A Brief Overview of Intelligent Data Analysis

Intelligent data analysis (IDA) or knowledge discovery in databases is defined in [23] as the "non-trivial process of identifying valid, novel, potentially useful and understandable (if not immediately, with some kind of further processing) patterns from the data". As it follows from this definition, in the IDA process, the data are the most important part of the discipline [23] and it is a complex process that includes the obtaining of the models and also other stages.

J.M. Cadenas (✉) · M.C. Garrido
Department of Information and Communications Engineering,
University of Murcia, Murcia, Spain
e-mail: jcadenas@um.es

M.C. Garrido
e-mail: carmengarrido@um.es

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_2

IDA is divided into the following stages, [23]:

- The "Data Integration and Collection" (selection) stage.
- The "Data Preprocessing" stage, related to the treatment of the data and the strategies that would be used to handle the available information.
- The "Data Mining" stage, related to the selection and application of the appropriate methods for the modeling from the available data and the obtaining of understandable models and high accuracy.
- The "Evaluation (interpretation) and Diffusion" stage.

Although all stages are fundamental to the development of the IDA process, its core is in the data preprocessing and data mining stages.

1.1 Data Preprocessing Stage

Data preprocessing can have a great impact on the performance of the data mining methods, [27]. One of the problems that must be faced in this stage is to understand and analyze the nature of the data avoiding the loss of useful information during the process. This stage includes, among others, the cleaning of data (such as the elimination of inconsistent data, treatment of missing values, etc.), data integration (multiple sources), data transformation (discretization, etc.) and reduction of data (attribute/instance selection) [27].

Specifically, the "discretization of continuous attributes" plays a critical role in IDA and has been studied in depth. Discretization consists in dividing the values of a numerical (continuous) attribute into a set of intervals. By means of the discretization, a numerical attribute can be more concise and easier to understand. In the general description of the discretization process, we can do the following taxonomy (there are other taxonomies for the different discretization methods such as that presented in [51]):

- Top-down methods: The attribute domains are progressively cut to construct a set of intervals.
- Bottom-up methods: They start with the individual values in the dataset that are fused progressively until constructing a set of intervals.

Among the top-down methods we find the ones proposed in [15, 33, 34, 46, 81]. Besides, the decision trees construction methods, such as, ID3 [71] and C4.5 [72], can be interpreted as top-down discretization methods. Among the bottom-up methods we find methods such as those proposed in [9, 44, 52]. All these methods generate classical discretization, i.e., crisp intervals.

Also, the "attribute selection" plays an important role in the IDA process and more specifically in the classification task. On the one hand the computational cost is reduced and on the other hand, a model is constructed from the simplified data and this improves the general abilities of classifiers. The first motivation is clear, since the computation time to build models is lower with a smaller number of attributes.

The second reason indicates that when the dimension is small, the risk of "overfitting" is reduced. Removing insignificant attributes of datasets can make the model more transparent and more comprehensible providing a better explanation of the system [53]. Therefore, the attribute selection addresses the problem of reducing dataset dimensionality by identifying an available attributes subset. Researchers have studied various aspects of attribute selection. One of the key aspects is to measure the goodness of an attribute subset determining an optimal one. Depending on evaluation criteria, attribute selection methods can be divided into the following categories, [29, 75]:

- Filter methods: These methods select subsets of attributes as a preprocessing step, independently of the chosen classifier.
- Wrapper methods: These methods use a method of data mining as a black-box to score attribute subsets according to their predictive power.
- Embedded methods: These methods select attributes in the training process and are usually specific to the given modeling method.
- Hybrid methods: These methods are a combination of filter and wrapper methods. Hybrid methods use the ranking information obtained using filter methods to guide the search in the optimization algorithms used by wrapper methods.

In literature we can find a variety of methods to carry out attribute selection, such as the proposed in [3, 42, 52, 75].

1.2 Data Mining Stage

The data mining (DM) stage is the more characteristic stage in the IDA process. The purpose of DM is the construction of models based on the data to produce new knowledge that can be used by the user. The model is a description of patterns and relationships in the data, which can be used to make predictions in a particular area, better understand the domain, improve performance or explain past situations. In practice, there are two types of models: Predictive (identify patterns to estimate future values using predictor attributes) and Descriptive (identify patterns that explain the data). In addition, different types of tasks are distinguished in DM. Each task has its own requirements and obtains a type of knowledge different from the obtained one by other tasks. Among the aimed tasks that obtain predictive models, we can find both the classification and the regression tasks; while clustering and association are tasks aimed at obtaining descriptive models. This stage includes the choice of the most appropriate task for the problem, the choice of the DM method, and finally the use and adaptation to the problem of the selected method, [27, 85].

We group these methods according to the type of model obtained. Without being exhaustive, we find models represented by discriminant functions, decision trees, neural networks, based on rules or based on instances.

- One of the most useful ways of representing a model is through a set of discriminant functions. The model in this case can be seen as a machine which computes c discriminant functions $g_i(x)$ and which selects for x the class ω_i with the highest value for the discriminant function [22, 26]. In this way the model is expressed as $g_i(x) = P(\omega_i/x)$, such that the maximum discriminant function is the maximum a posteriori probability. When the discriminant functions are linear functions we find methods such as descending gradient, Newton's algorithm, the Perceptron criterion [32, 36]. When the discriminant functions are complex density functions, these can be approximated by a mixture of simpler density functions.
- The models based on instances approximate an unknown density function using an averaged version of the density based on the probability of a specific vector's falling within a certain region of the attribute space [22]. The methods based on these models have no learning phase since the model is formed by the dataset instances. There are two common methods based on these models: Parzen method and k neighbors method [21, 31, 58].
- The methods which model the problem through decision trees are useful for finding structures in high dimensionality spaces or when the conditional densities of the classes are unknown or are multimodal. Some basic and well-known methods to generate decision/regression trees are ID3 [71], C4.5 [72] and CART [10].
- Rules based methods model a system through a base of rules (if-then) constructed from the instances. Some methods for obtaining rules (association rules) are based on the concept of frequent items sets and use counting and minimum support methods [79]. Other methods obtain rules covering the instances (cover methods) such as those based on CN2 [17] and AQ algorithms [57]. Genetic algorithms/programming [51, 86] have also been used to generate rules.
- Other type of model is the neural network. Neural networks are a very powerful computation paradigm allowing complex problems with possible non linear interactions between the attributes. Among the most important neural networks we can find the multilayer Perceptron which generates more than one boundary of separating in the attributes space [32, 36, 74].
- There is a further group of methods whose aim is to generate groupings of data and these are known as clustering. The aim of cluster analysis is to find a valid and convenient organization for the data and an underlying structure. Within these methods we can include Kohonen's self-organizing maps [45], those based on the K-means algorithms which obtain partitional cluster [62], in contrast to the hierarchical methods which do not establish a priori the groups number [25].

2 Intelligent Data Analysis and Soft Computing

In [56] several paradigms introduced with the data analysis are identified. Among them, the management and processing of data respecting the true nature of them (imperfect data) are included. Therefore, by focusing on the data, and before applying any stage of the IDA process, we must take into account the nature of these data to ensure the success of the process. This means that depending on the nature and precision of these data, we must apply different methods depending on their degree of tolerance to them. A clear example to illustrate the problem of the different nature of the data and the importance of tolerance to different types of imperfect data is the problem of parking a car [91], where most of the population is able to do it easily. Therefore, we need methods that can extract knowledge and handle imperfect data, in order to provide quality information and generate useful knowledge.

Generally, the IDA process uses and combines different methods and tools from a variety of disciplines [5]. Due to possible imperfection in the data, tools provided by the Fuzzy Sets theory [90] and, in general, Soft Computing (SC) [7, 82, 91] are quite adequate. In this way, the hybridization of the SC methods with IDA is a field that is gaining more importance. The methods proposed by SC and their applications have been very important in recent years, and in particular, the advances in the hybridization of SC with IDA are aimed at obtaining more flexible methods with results more efficient compared to the classical methods [30, 61]. In this framework, we comment on different methods proposed from two levels: In a first level, if the SC elements are incorporated in the design of methods/models; and, in a second level, if they are incorporated for the treatment of imperfect information, additionally.

2.1 Data Preprocessing in Soft Computing Framework

In the data preprocessing stage, SC has generally been applied to the design of flexible methods for the different tasks of this stage. Although most of them use SC in their development, to our knowledge, the methods that allow and management imperfect data are seldom studied.

In particular, in the discretization of numerical attribute we find methods that allow the use of membership degree to intervals (denoted by fuzzy discretization methods). These methods are grouped according to the used algorithm.

- Decision tree based methods: In [40, 43, 63] different approaches for the fuzzy discretization of numerical attributes are proposed. All of them use a fuzzy decision tree combined with some basic strategy.
- Clustering based methods: These methods are based on dividing a numerical attribute domain into fuzzy partitions by using fuzzy clustering. In particular, several methods using the fuzzy c-means method are proposed in [59, 64, 80].
- Genetic algorithm based methods: The genetic algorithms (GA) are combined with existing specialized methods to create hybrid algorithm that improve the overall results. In particular, we can find several methods, [16, 18], using strategies of classical/fuzzy discretization together a genetic algorithm to optimize the number of partitions, interval limits and the degree of overlaps of these limits.
- Hybrid methods: In the literature we can also find methods based on combinations of two or more methods. In [88] a cluster and a neural network (NN) are used. In

[76] the combination of the FCM clustering algorithm and a GA are used, and in [48, 73, 84] a kd-tree and a minimum spanning tree are used.

In attribute selection, there are a lot of methods using SC in their development but they perform the selection from crisp data.

- Attribute selection methods using SC for their design can be find in [3, 16, 42] where a neural network, a GA or an ant colony (AC) are used, respectively. There are other methods that also use elements of the fuzzy set theory as in [53, 83] where a fuzzy criteria or fuzzy entropy are used, or in [87] where the attribute selection is performed using the fuzzy evidence theory.
- To perform the attribute selection from imperfect data we can find several proposals: in [41] a method taking into account the uncertainty in the data through fuzzy-rough sets is presented. This method employs fuzzy-rough sets to provide a means by which discrete or real-valued noisy data (or a mixture of both) can be effectively reduced without the need for user-supplied information. In [77, 78] a fuzzy mutual information measure between two fuzzified numerical attributes to handle imprecise data is used (they define a new extended version of Battiti's filter attribute selection method). This measure is used in combination with a genetic optimization to define the method proposed.

Table 1 shows the summary of papers discussed.

Table 1 Hybridization of data preprocessing with Soft Computing: summary of papers

Method based on …		SC at method level	SC at minable view level	
				Allowed data
Fuzzy discretization	Fuzzy decision trees	[40, 43, 63]	–	–
	Fuzzy clustering	[59, 64, 80]	–	–
	GA to optimize	[16, 18]	–	–
	kd tree—spanning tree	[48, 73, 84]	–	–
	Cluster—GA	[76]	–	–
	Cluster—NN	[88]	–	–
Attribute selection	NN, GA, AC	[3, 16, 42]	[77, 78]	Fuzzy sets
	Fuzzy criteria/entropy	[53, 83]	–	–
	Fuzzy evidence theory	[87]	–	–
	Fuzzy-rough metric	–	[41]	Fuzzy-rough sets

2.2 Data Mining in Soft Computing Framework

SC has also been applied in the DM stage, and, to our knowledge, the methods that allow and management imperfect data are seldom studied. From this, we can consider the DM methods hybridized with SC in two levels:

- At the level of generated models: Methods that generate models described in the framework of SC. These models are more interpretable and we can find elements of SC in rule-based systems, methods based on k-nearest neighbors, decision trees, clustering and support vector machines.
  In 1971 Zadeh proposed the design of rules if-then using linguistic variables that can be provided by a group of experts or obtained through DM methods. So, among others, in [4] a set of fuzzy rules is obtained using a method based on genetic programming, in [24] a set of fuzzy rules is obtained in unbalanced problems using a genetic selection process of rules, in [37] different weights are assigned to a set of fuzzy rules using heuristic methods, and, in [65] an initial set of fuzzy rules is constructed by clustering and then are optimized using a neuro-fuzzy learning algorithm.
  Among the fuzzy versions of the k-nearest neighbors rule we can highlight works that assign memberships degree of each instance to each class, use fuzzy distance measures, use different ways of combining the votes of neighbors, etc. A complete review of these methods is carried out in [20].
  Also, fuzzy decision trees have been designed as the proposed in [66] that obtains the best fuzzy partition of the best attribute in each node to split. Using fuzzy decision trees, fuzzy ensembles are proposed as in [19] where an ensemble is constructed from a non-fuzzy tree construction algorithm that subsequently is transformed to fuzzy.
  With the aim to construct data partitions that allow an instance belongs to more than one partition, fuzzy clustering algorithms have been developed such as the fuzzy C-means proposed in [6]. Different versions of this algorithm are found in [35] to extend it to nominal data, in [49] to deal with missing values through intervals or in [80] to deal with fuzzy values.
  Also, fuzzy versions of support vector machines have been designed. So, in [50] a membership degree to each class is assigned to each instance, allowing that each one contributes in a different way in the learning of the decision surface. In [1] a method for multilabel classification is generalized. For each multilabel class, a region with the associated membership function is defined and an instance is classified into a multilabel class whose membership function is the largest.
- At minable view level: Methods that besides incorporating the SC elements, support input imperfect data. In this case, the methods allow us that the data are composed of attributes described by imperfect values. This generates the following advantages: (1) methods can interpret the imprecision/uncertainty expressed in the data and generate robust models to these types of information without transforming the true nature of them; (2) data preprocessing is simplified by not carrying out these transformations (replacement, deleting data, …); and (3) the minable view

contains a greater number of instances because the imprecise and uncertain data are not discarded. In general, significant efforts are being carried out to incorporate the treatment of imperfect data into DM methods using SC.

Thus we can find works that incorporate the treatment of fuzzy values. There are fuzzy decision trees based on a fuzzy partition of numerical attributes. This partition is used in the test of nodes as in [38, 47]. Fuzzy partitions of numerical attributes are also use in the construction of fuzzy ensembles to incorporate fuzzy values. This approach is used in [39] where to select the test of each node, the set of the best attributes for partitioning that node is used or in [55] where a fuzzy ensemble for each class value of the problem is constructed. In [60] a fuzzy version of multilayer perceptron is presented which performs the learning from fuzzy values. In [68] a genetic classifier based on fuzzy rules is obtained from data described with fuzzy values. In [69, 70] Adaboost and FURIA algorithms are extended in order to obtain fuzzy rules from this type of values. In [67] an algorithm to obtain a set of fuzzy association rules from a fuzzy partition is proposed. As particular cases of fuzzy values, some works deal with values expressed by intervals as in [47, 67–70].

On the other hand, the set of methods that allow the existence of missing values is considerable. We highlight only a few that allow the treatment of some other type of imperfect information as [38, 39] or as in [47], where missing values are only allowed in the classification phase.

Finally, there is a considerable set of methods that have considered the possibility that an instance has more than one associated class value (multi-valued class), but few extend this possibility to other nominal attributes of a problem (multi-valued attributes). So, among the first we can find works as [68] where class may be defined by a crisp set, or [89] where a fuzzy k-nearest neighbor method is used to allow that an instance can belong to more than one class with several degrees. In [54] we can find a comparison of this kind of methods.

Table 2 shows the summary of papers discussed.

Table 2 Hybridization of data mining with Soft Computing: summary of papers

Method based on …	SC at method level	SC at minable view level	
			Allowed data
Fuzzy rules	[4, 24, 37, 65]	[67–70]	Fuzzy sets, intervals
	–	[68]	Fuzzy sets, intervals, multivalued class
k-nearest neighbors	[20]	[89]	Multivalued class
Fuzzy decision trees	[19, 66]	[55]	Fuzzy sets
		[38, 39]	Fuzzy sets, missing
		[47]	Fuzzy sets, intervals, missing
Fuzzy clustering	[6, 35, 49, 80]	[35, 49, 80]	Nominal, fuzzy sets, intervals
Support vector m	[1, 50]	[1]	Multivalued class
Neural network	–	[60]	Fuzzy sets

3 Hybridization on the Two Level of Soft Computing and Data Preprocessing/Mining Methods

In this section we describe the characteristic elements of two methods in the data preprocessing stage and three methods in DM stage that use SC in the two levels commented: at model/technique level and at minable view level. Due to the high flexibility in the design of these methods, they can easily be extended to support new types of imperfect data.

A more detailed analysis of these methods can be found in papers [11, 13] for the preprocessing methods and papers [8, 12, 14, 28] for the DM ones.

3.1 Notation, Types and Representation of Imperfect Values

Let us consider a set of instances E, where each instance $\mathbf{x}$ is characterized by n attributes in a vector $(x_1, x_2, \ldots, x_n)$ (the n-th attribute represents the class). The domains of each attribute, $\Omega_{x_1}, \Omega_{x_2}, \ldots, \Omega_{x_{n-1}}$, can be numerical or nominal, while the domain of the class Ω_{x_n} (nominal attribute) can take the values $\{\omega_1, \omega_2, \ldots, \omega_I\}$.

The numerical attributes are represented by fuzzy sets with a trapezoidal fuzzy membership function [2] $\mu(x)$ defined by a quadruple (a, b, c, d):

$$\mu(x) = \begin{cases} 0 & x < a \text{ or } x \geq d \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & b \leq x < c \\ \frac{d-x}{d-c} & c \leq x < d \end{cases}$$

With this representation, the methods use the following values:

- Crisp values are represented by the quadruple (a, a, a, a).
- Interval values $[a, b]$ are represented by the quadruple (a, a, b, b).
- Fuzzy values are represented by trapezoidal fuzzy membership functions.
- Missing values include pieces of information that are unknown. These values are represented by the quadruple $(min_i, min_i, max_i, max_i)$, where min_i and max_i are, respectively, the minimum and maximum values of Ω_{x_i} included in the dataset.

The nominal attributes (including the class attribute) are represented by fuzzy subsets $\{\mu(h_1)/h_1, \ldots, \mu(h_s)/h_s\}$, where h_j is a value into attribute domain and $\exists h_k \in \Omega_i : \mu(h_k) = 1$. With this representation, the methods use the following values:

- Crisp values are represented by the fuzzy subset $\{1/h_j\}$.
- Crisp subset values consider more than a possible nominal value. They are represented as $\{1/h_1, \ldots, 1/h_s\}$.

- Fuzzy subset values consider more than one nominal value with a membership value $\mu \in [0, 1]$. They are represented using the notation introduced above.
- Missing nominal values are represented using a fuzzy subset that contains all possible values with membership degree equals to 1.

3.2 OFP_CLASS: A Hybrid Method for Attribute Discretization

In [13], OFP_CLASS method is proposed to data preprocessing. It is a hybrid method for discretizing numerical (continuous) attributes by means of fuzzy sets, which constitute a fuzzy partition of the domains of these attributes. The aim of this method is to find an attribute partition so that the fuzzy classification methods obtain better results. The OFP_CLASS method can deal with datasets with imperfect values and it is labeled as supervised, local, top down, and incremental, using the entropy as measure to obtain the partition.

The OFP_CLASS method is composed of two stages (Fig. 1): (a) In the first stage, crisp intervals are defined for each attribute using a fuzzy decision tree (FDT); and (b) in the second stage, these intervals are used as the starting point to form an optimal fuzzy partition for classification. In this second stage, a genetic algorithm is used to determine the cardinality and fuzzy boundary of these intervals.

The partition obtained for each attribute guarantees:

- Completeness (no point in the domain is outside the fuzzy partition), and
- Strong fuzzy partition (it verifies that $\forall x \in \Omega_i$, $\sum_{f=1}^{F_i} \mu_{B_f}(x) = 1$, where $B_1, \ldots, B_{F_i}$ are the F_i fuzzy sets for the partition corresponding to the i-th numerical attribute with Ω_i domain).

The FDT used in the first stage allows the dealing of imperfect data, and for this, uses a specific information gain, G_i, for each attribute i in order to choose the best attribute to divide a node. Function G_i uses the standard information associated with the node (taking into account the membership degree of an instance to the node and

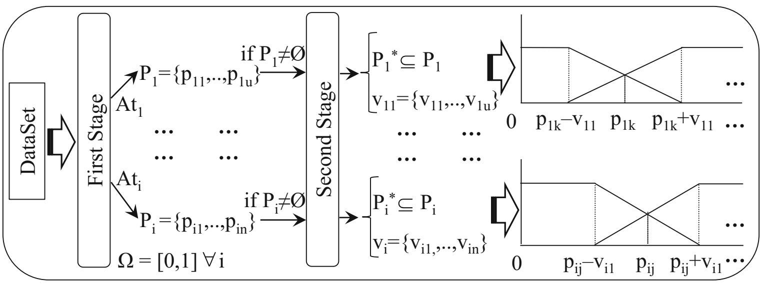


Fig. 1 Scheme for the discretization of numerical attributes using the OFP_CLASS method

the membership degree of example e_j to each class) and a factor which represents the standard information obtained by dividing the node using attribute i adjusted to the existence of missing values. We must highlight that, in the second stage, the fitness function of the genetic algorithm is defined by $\frac{\sum_{i=1}^{n} I_i}{\sum_{i=1}^{n} H_i}$ where I_i and H_i are the information gain and entropy of attribute i respectively, taking into account the crisp intervals obtained in the first stage.

OFP_CLASS method is an effective strategy and it obtains very good results when is compared with other methods of the literature. These results have been validated by applying statistical techniques to analyze the behavior of different methods in each experiment.

3.3 FRF_fs: A Filter-Wrapper Method for Attribute Selection

In [11] is proposed the FRF_fs method of attribute selection to data preprocessing which can handle imperfect data. This method is based on a Fuzzy Random Forest ensemble (a method that supports imperfect data, [8, 12]) and is classified as a Filter-Wrapper method with sequential forward selection on the subset of attributes obtained by the Filter method and using a ranking obtained with these attributes. This method consists of the following main steps (Fig. 2): (1) Scaling and discretization process of the attribute set; and attribute pre-selection using the discretization process (Filter); (2) Ranking process of the attribute pre-selection; and (3) Wrapper attribute selection based on cross-validation.

Note that in each step the approach obtains information useful to the user (pre-selected attribute subset, pre-selected attribute subset ranking and optimal attribute subset). Some details of these steps are discussed below.

- Filter method for attribute pre-selection
  Initially, the method carry out a scaling and discretization (in [13], a hybrid method for the fuzzy discretization of numerical attributes is presented), and as in the discretization process some attributes may be discretized into a single interval,

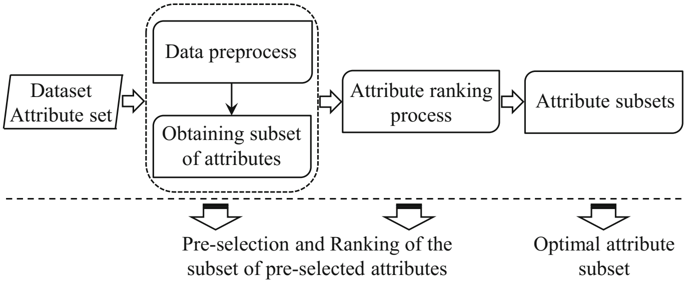


Fig. 2 Framework for the attribute selection using the FRF_fs method

these latter attributes can be removed. Thus, the method obtain a pre-selection of the attribute set.

- Attribute importance (Ranking process)
  From the pre-selected attribute subset and through a fuzzy random forest ensemble, the method obtains a vector *RANK* ordered, in descending order, of this attribute subset. This vector is obtained from the value of each attribute x_i as $RANK = \sum_{t=1}^{T} W \cdot IMP_t$, where the information provided by the T trees of the fuzzy random forest ensemble is aggregated using an OWA operator. Values IMP_t are obtained from the information gain of nodes in the FDT t to each attribute x_i, and from the accuracy of FDT t classifying the *OOB* dataset.
- Wrapper for attribute final selection
  Once the ranking of the pre-selected attribute subset, *RANK*, is obtained, the method find an optimal subset of attributes. The process adds a single attribute at a time following the *RANK* vector. The several attribute subsets obtained by this process are evaluated by a method that supports imperfect data using a cross-validation. In particular, and using a fuzzy random forest ensemble, an ascending sequence of fuzzy random forest models is constructed, by invoking and testing the stepwise attributes.

The efficiency and effectiveness of the FRF_fs method is proved through several experiments using both high dimensional and imperfect datasets. The method shows a good performance (not only classification accuracy, but also with respect to the number of selected attributes) and good behavior both with high dimensional datasets and with imperfect datasets.

3.4 EMFGN: A Method Based on Gaussian Mixture Models

Extended Mixture of Factorized Generalized Normal (EMFGN) method [28] is a predictive DM method for performing learning and inference from imperfect data. The method obtains an explicit expression of the model-observation joint function of the attributes, where both the model expression and the input information are interpreted and represented in a common framework. The Dempster-Shafer Evidence Theory (DSET) is the framework that allows its interpretation as mass functions defined on the domains of the single attributes.

In Fig. 3, the general scheme of the process followed by EMFGN method is shown. From the dataset with imperfect information, the method provides a model reflecting the joint dependence of the attributes by means of a mixture of factorized normals. This model and the input available information are interpreted and represented in the DSET in order to combine them (using the Dempster-Shafer's combination rule). The model provided by EMFGN method is the following:

$$p(z) = \sum_{ir} P\{C_i\}\pi_r \prod_{j=1}^{n} F_{irj}(m_{rj}(\Theta_j) \oplus m_{ij}(\Theta_j))$$

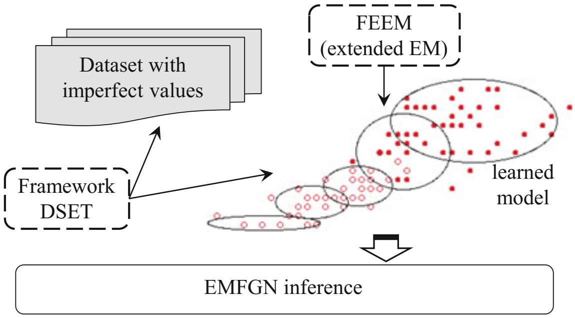


Fig. 3 A general scheme of the EMFGN method

where:

- $\Theta_j \in \mathscr{P}(\Omega_{z_j})$ and $\mathscr{P}(\Omega_{z_j})$ is the set of parts of Ω_{z_j}.
- $m_{rj}(\Theta_j)$ is the likelihood function of the r-th component of the input information expressed through a mass function.
- $m_{ij}(\Theta_j)$ is the mass function corresponding to the i-th component of the model.
- F_{irj} is a necessary normalization factor in the combination of two mass functions.
- $P\{C_i\}\pi_r$ is the product of the likelihood function of input information in its r-th component and the expression of the model in the i-th component.

In this framework, the EMFGN method uses the FEEM algorithm in the learning phase. This algorithm is an extended EM algorithm to allow both the imperfect information and the model represented in DSET.

From the learned model, EMFGN method can infer both nominal and numerical attributes. To numerical attributes, the method infers the value $z_j = \sum_{ir} \alpha_{ir}\bar{m}_{irj}$, and to nominal attributes, the method infers the value $z_j = argmax_w \sum_{ir} \alpha_{ir} m_{irj}(\omega)$, with $\omega \in \Omega_{z_j}$. The value α_{ir} indicates the likelihood of the r-th component of the input information having been generated by the i-th component of the mixture. $m_{irj}(\cdot)$ is a mass function combining the input information and the model to the attribute j, and the value $\bar{m}_{irj}$ is the average value of $m_{irj}(\Theta_j)$.

The results obtained are very satisfactory with the advantage of having a global model to be able to perform inference on any attribute of an instance.

3.5 FRF: A Method Based on an Ensemble of Fuzzy Decision Trees

Fuzzy Random Forest (FRF) method [8, 12] is a multiple classifier system (ensemble) to DM. FRF is a predictive method for classification and show us its ability to handle imperfect data both in the model learning and in the inference process.

In Fig. 4, the general scheme of the process followed by FRF method is shown. FRF obtains a model with the structure of an ensemble based on FDTs. The learning

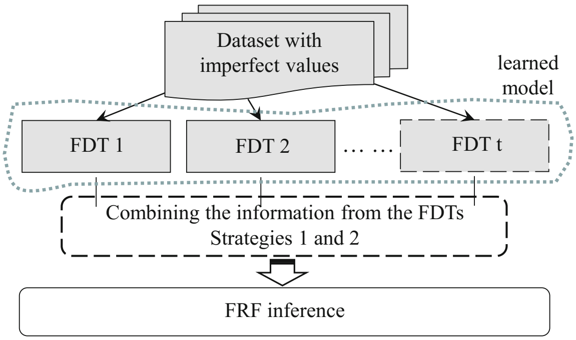


Fig. 4 A general scheme of the FRF method

phase generates FDTs with the following characteristics: (a) each FDT is constructed from a dataset obtained by bagging, (b) the FDTs are constructed without considering all the attributes to split the nodes (a random subset of the set of attributes available at each node is selected), (c) the numerical attributes are discretized by fuzzy partitions, (d) each FDT is constructed to the maximum size and without pruning, (e) a function ($\chi_{t,N}(\cdot)$) is used to indicate the degree with which an instance satisfies the conditions that lead to node N of tree t, and (f) FDTs support instances with imperfect values (a function $\mu_{simil}(\cdot)$ is used to measure the membership degree of these types of values to the fuzzy sets forming the partition of the numerical attributes).

From the obtained model, FRF method uses two strategies to combine the information of several FDTs and to obtain the final decision for a target instance. Strategy 1 combines the information from the different leaves reached in each FDT to obtain the decision of each individual FDT and then applying the same or another combination method to generate the global decision of the FRF model. Strategy 2 combines the information from all reached leaves from all FDTs to generate the global decision of the FRF model.

The method assigns class ω_M to a new instance such that $\omega_M = argmax_i\{D\_FRF_i\}$ where D_FRF is a vector with size I that indicates the confidence assigned by the method to each class i. The vector elements are obtained from the support for each class in the leaves reached when applying the several strategies and combination methods.

The results obtained by FRF method are promising concluding that by using imperfect values instead of crisp, we capture better the nature of the underlying information.

3.6 KNN$_{imp}$: A Method Based on Instances

The kNN$_{imp}$ method [14] is a k-nearest neighbors classifier from datasets with imperfect values to DM. Figure 5 shows the general scheme followed by kNN$_{imp}$ method. This method belongs to the methods with lazy learning, that is, the method does not

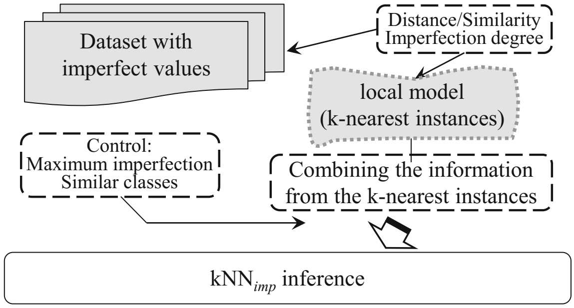


Fig. 5 A general scheme of the kNN$_{imp}$ method

need of an explicit learning phase. Therefore, this method requires that all dataset instances are stored.

To classify a instance, the kNN$_{imp}$ method computes its "k" nearest instances and generates a class value from them (a local model dependent on the new instance has been constructed). By containing imperfect values the dataset, the importance of each instance (neighbor) in the output decision is based on relative distance/similarity $d_{imp}(\cdot,\cdot)$ (distance/similarity measures that support imperfect data) and its degree of imperfection. Specifically, for each instance, two weights are calculated depending on its degree of imperfection $p(\cdot)$ and its distance/similarity $q(\cdot)$.

Furthermore, the overall degree of imperfection in "k" nearest instances is measured, if it is too high, the classification is not performed. To establish the maximum degree of imperfection, kNN$_{imp}$ method uses the parameter U_I.

Once the local model is obtained (k nearest instances), kNN$_{imp}$ method combines the information provided for each neighbor instance (weights $p(\cdot)$ and $q(\cdot)$) to obtain the set of possible weighted classes. The class with the highest score is chosen as output, together with other classes whose score is similar to the highest. To assess if a class should be included in the final output, this method uses the threshold U_D.

The method obtains a fuzzy subset $\{\mu(\omega_i)/\omega_i\}$ as possible values to the class attribute of the new instance where $\mu(\omega_i) = \frac{\sum_j^k \mu^j(\omega_i)p(x^j)q(x^j)}{\sum_j^k \sum_i \mu^j(\omega_i)p(x^j)q(x^j)}$ and $\mu^j(\omega_i)$ is the membership degree of the j-th neighbor to the class value ω_i. Therefore, the method assigns to the new instance the class $\omega_M = argmax_i\{\mu(\omega_i)\}$ or the fuzzy subset $\{\omega_M, \omega_t\}$, with $\frac{\omega_M - \omega_t}{\omega_M} > U_D$.

The kNN$_{imp}$ classifier is robust when working with imperfect data and maintains a good performance when is compared with other methods in the literature, applied to datasets with or without imperfection.

4 Conclusions

In data-driven application domains, the suitable use of available information is very important. Because of this, it becomes increasingly necessary to design methods that support different types of information (imperfect or not) and obtain more flexible models with an appropriate behavior. In this framework, the hybridization of the tools provided by Soft Computing and Intelligent Data Analysis methods is a field that is gaining more importance. In this work, some proposals that carry out this hybridization obtaining quite satisfactory results are commented and analyzed. For this reason we consider that it is a field in which new proposals must be made with the objective of approaching the Intelligent Data Analysis process from datasets that express the true nature of the information.

Acknowledgements Supported by the projects TIN2014-52099-R (EDISON) and TIN2014-56381-REDT (LODISCO) granted by the Ministry of Economy and Competitiveness of Spain (including ERDF support).

References

1. Abe, S.: Fuzzy support vector machines for multilabel classification. Pattern Recogn. **48**(6), 2110–2117 (2015)
2. Barua, A., Mudunuri, L.S., Kosheleva, O.: Why trapezoidal and triangular membership functions work so well: Towards a theoretical explanation. J. Uncertain Syst. **8**(3), 164–168 (2014)
3. Battiti, R.: Using mutual information for selection features in supervised neural net learning. IEEE Trans. Neural Netw. **5**, 537–550 (1994)
4. Berlanga, F.J., Rivera, A., del Jesús, M.J., Herrera, F.: GP-COACH: genetic programming-based learning of compact and accurate fuzzy rule-based classification systems for high-dimensional problems. Inf. Sci. **180**(8), 1183–1200 (2010)
5. Berthold, M.R., Borgelt, C., Höppner, F., Klawonn, F.: Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data. Springer (2010)
6. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press (1981)
7. Bonissone, P.: Soft computing: the convergence of emerging reasoning technologies. Soft Comput. **1**(1), 6–18 (1997)
8. Bonissone, P., Cadenas, J.M., Garrido, M.C., Díaz-Valladares, R.A.: A fuzzy random forest. Int. J. Approximate Reasoning **51**(7), 729–747 (2010)
9. Boulle, M.: Khiops: a statistical discretization method of continuous attributes. Mach. Learn. **55**(1), 53–69 (2004)
10. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. CRC Press (1984)
11. Cadenas, J.M., Garrido, M.C., Martínez, R.: Feature subset selection filter-wrapper based on low quality data. Expert Syst. Appl. **40**(16), 6241–6252 (2013)
12. Cadenas, J.M., Garrido, M.C., Martínez, R., Bonissone, P.P.: Extending information processing in a fuzzy random forest ensemble. Soft Comput. **16**(5), 845–861 (2012)
13. Cadenas, J.M., Garrido, M.C., Martínez, R., Bonissone, P.P.: OFP_CLASS: a hybrid method to generate optimized fuzzy partitions for classification. Soft Comput. **16**(4), 667–682 (2012)
14. Cadenas, J.M., Garrido, M.C., Martínez, R., Muñoz, E., Bonissone, P.P.: A fuzzy k-nearest neighbor classifier to deal with imperfect data. Soft Comput. (2017). doi:10.1007/s00500-017-2567-x
15. Chan, C.C., Batur, C., Srinivasan, A.: Determination of quantization intervals in rule based model for dynamic systems. In: Proceedings of the International Conference on Systems, Man, and Cybernetics, pp. 1719–1723 (1991)
16. Choi, Y.S., Moon, B.R.: Feature selection in genetic fuzzy discretization for the pattern classification problems. IEICE Trans. Inf. Syst. **90**(7), 1047–1054 (2007)

17. Clark, P., Boswell, R.: Rule induction with CN2: Some recent improvements. In: Y. Kodratoff (ed.) Lecture Notes in Artificial Intelligence. Machine Learning EWSL-91, pp. 151–163. Springer (1991)
18. Cox, E., Taber, R., O'Hagen, M.: The Fuzzy Systems Handbook, 2nd edn. Academic Press (1998)
19. Crockett, K., Bandar, Z., Mclean, D.: Growing a fuzzy decision forest. In: Proceedings of the 10th IEEE International Conference on Fuzzy Systems, vol. 2, pp. 614–617 (2001)
20. Derrac, J., García, S., Herrera, F.: Fuzzy nearest neighbor algorithms: taxonomy, experimental analysis and prospects. Inf. Sci. **260**, 98–119 (2014)
21. Djouadi, A., Bouktache, E.: A fast algorithm for the nearest-neighbor classifier. IEEE Trans. Pattern Anal. Mach. Intell. **19**(3), 277–282 (1997)
22. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley (2001)
23. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery: an overview. In: U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (eds.) Advances in Knowledge Discovery and Data Mining, pp. 1–34. American Association for Artificial Intelligence (1996)
24. Fernández, A., del Jesus, M.J., Herrera, F.: Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets. Int. J. Approximate Reasoning **50**(3), 561–577 (2009)
25. Fisher, D.H.: Knowledge acquisition via incremental conceptual clustering. Mach. Learn. **2**(2), 139–172 (1987)
26. Fukunaga, K.: Introduction to statistical pattern recognition. Academic Press (1990)
27. García, S., Luengo, J., Herrera, F.: Data Preprocessing in Data Mining. Springer (2015)
28. Garrido, M.C., Cadenas, J.M., Bonissone, P.P.: A classification and regression technique to handle heterogeneous and imperfect information. Soft Comput. **14**(11), 1165–1185 (2010)
29. George, G., Raj, V.C.: Review on feature selection techniques and the impact of svm for cancer classification using gene expression profile. Int. J. Comput. Sci. Eng. Surv. **2**(3), 16–27 (2011)
30. Ghosh, A., Pal, S.K.: Soft computing approach to pattern recognition and image processing. World Scientific (2002)
31. Hamamoto, Y., Uchimura, S., Tomita, S.: A bootstrap technique for nearest neighbor classifier design. IEEE Trans. Pattern Anal. Mach. Intell. **19**(1), 73–79 (1997)
32. Hecht-Nielsen, R.: Neurocomputing. Addison-Wesley (1990)
33. Ho, K.M., Scott, P.D.: Zeta: a global method for discretization of cotitinuous variables. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, pp. 191–194 (1997)
34. Holte, R.C.: Very simple classification rules perform well on most on most commonly used datasets. Mach. Learn. **11**, 63–90 (1993)
35. Huang, Z., Ng, M.K.: A fuzzy k-modes algorithm for clustering categorical data. IEEE Trans. Fuzzy Syst. **7**(4), 446–452 (1999)
36. Isasi, P., Galván, I.: Las redes neuronales artificiales - Un enfoque práctico. Pearson - Prentice Hall (2004)
37. Ishibuchi, H., Yamamoto, T.: Rule weight specification in fuzzy rule-based classification systems. IEEE Trans. Fuzzy Syst. **13**(4), 428–435 (2005)
38. Janikow, C.Z.: Fuzzy decision trees: issues and methods. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **28**(1), 1–14 (1998)
39. Janikow, C.Z.: Fuzzy decision forest. In: Proceedings of the 22nd International Conference of the North American Fuzzy Information Processing Society, pp. 480–483 (2003)
40. Janikow, C.Z., Fajfer, M.: Fuzzy partitioning with fid3. 1. In: Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society, pp. 467–471 (1999)
41. Jensen, R., Shen, Q.: Fuzzy-rough sets assisted attribute selection. IEEE Trans. Fuzzy Syst. **15**(1), 73–89 (2007)
42. Kabir, M., Shahjahan, M., Murase, K.: A new hybrid ant colony optimization algorithm for feature selection. Expert Syst. Appl. **39**, 3747–3763 (2012)

43. Kbir, M.A., Benkirane, H., Maalmi, K., Benslimane, R.: Hierarchical fuzzy partition for pattern classification with fuzzy if-then rules. Pattern Recogn. Lett. **21**(6), 503–509 (2000)
44. Kerber, R.: ChiMerge: discretization of numeric attributes. In: Proceedings of the Tenth National Conference on Artificial Intelligence, pp. 123–128 (1992)
45. Kohonen, T.: Self-Organising Maps, 3rd edn. Springer (2001)
46. Kurgan, L.A., Cios, K.J.: Caim discretization algorithm. IEEE Trans. Knowl. Data. Eng. **16**(2), 145–153 (2004)
47. Lee, K.M., Lee, K.M., Lee, J.H., Lee-Kwang, H.: A fuzzy decision tree induction method for fuzzy data. Proceedings of the IEEE International Conference on Fuzzy Systems **1**, 16–21 (1999)
48. Li, C.: A combination scheme for fuzzy partitions based on fuzzy majority voting rule. In: Proc. of the International Conference on Networks Security, Wireless Communications and Trusted Computing, vol. 2, pp. 675–678 (2009)
49. Li, D., Gu, H., Zhang, L.: A fuzzy c-means clustering algorithm based on nearest-neighbor intervals for incomplete data. Expert Syst. Appl. **37**(10), 6942–6947 (2010)
50. Lin, C.F., Wang, S.D.: Fuzzy support vector machines. IEEE Trans. Neural Netw. **13**(2), 464–471 (2002)
51. Liu, H., Hussain, F., Tan, C.L., Dash, M.: Discretization: an enabling technique. Data Min. Knowl. Discov. **6**(4), 393–423 (2002)
52. Liu, H., Setiono, R.: Feature selection via discretization. IEEE Trans. Knowl. Data. Eng. **9**(4), 642–645 (1997)
53. Luukka, P.: Feature selection using fuzzy entropy measures with similarity classifier. Expert Syst. Appl. **38**(4), 4600–4607 (2011)
54. Madjarov, G., Kocev, D., Gjorgjevikj, D., Džeroski, S.: An extensive experimental comparison of methods for multi-label learning. Pattern Recogn. **45**(9), 3084–3104 (2012)
55. Marsala, C.: Data mining with ensembles of fuzzy decision trees. In: Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining, pp. 348–354 (2009)
56. Mayer-Schönberger, V., Cukier, K.: Big Data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt (2013)
57. Michalski, R.S., Mozetic, I., Hong, J., Lavrac, N.: The AQ15 inductive learning system: an overview and experiments. In: Proceedings of the International Meeting on Advances in Learning (1986)
58. Micó, M.L., Oncina, J., Vidal, E.: A new version of the nearest-neighbour approximating and eliminating search algorithm with linear preprocessing time and memory requirements. Pattern Recogn. Lett. **15**(1), 9–17 (1994)
59. Mirkin, B.: Mathematical Classification and Clustering. Kluwer Academic Publishers (1996)
60. Mitra, S., Pal, S.K.: Fuzzy multi-layer perceptron, inferencing and rule generation. IEEE Trans. Neural Netw. **6**(1), 51–63 (1995)
61. Mitra, S., Pal, S.K., Mitra, P.: Data mining in soft computing framework: a survey. IEEE Trans. Neural Netw. **13**(1), 3–14 (2002)
62. Moody, J., Darken, C.J.: Fast learning in networks of locally-tuned processing units. Neural Comput. **1**(2), 281–294 (1989)
63. Myles, A.J., Brown, S.D.: Induction of decision trees using fuzzy partitions. J. Chemom. **17**(10), 531–536 (2003)
64. Nascimento, S., Mirkin, B., Moura-Pires, F.: A fuzzy clustering model of data and fuzzy c-means. Proceedings of the Ninth IEEE International Conference on Fuzzy Systems **1**, 302–307 (2000)
65. Nauck, D., Kruse, R.: A neuro-fuzzy method to learn fuzzy classification rules from data. Fuzzy Sets Syst. **89**(3), 277–288 (1997)
66. Olaru, C., Wehenkel, L.: A complete fuzzy decision tree technique. Fuzzy Sets Syst. **138**(2), 221–254 (2003)
67. Palacios, A.M., Palacios, J.L., Sánchez, L., Alcalá-Fdez, J.: Genetic learning of the membership functions for mining fuzzy association rules from low quality data. Inf. Sci. **295**, 358–378 (2015)

68. Palacios, A.M., Sánchez, L., Couso, I.: Extending a simple genetic cooperative-competitive learning fuzzy classifier to low quality datasets. Evol. Intell. **2**(1–2), 73–84 (2009)
69. Palacios, A.M., Sanchez, L., Couso, I.: Boosting of fuzzy rules with low quality data. Multiple-Valued Logic Soft Comput. **19**(5–6), 591–619 (2012)
70. Palacios, A.M., Sanchez, L., Couso, I.: An extension of the FURIA classification algorithm to low quality data. In: Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, pp. 679–688 (2013)
71. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**, 81–106 (1986)
72. Quinlan, J.R.: C4.5 Programs for Machine Learning. Morgan Kaufmann (1993)
73. Redmond, S.J., Heneghan, C.: A method for initialising the k-means clustering algorithm using kd-trees. Pattern Recogn. Lett. **28**(8), 965–973 (2007)
74. Ruck, D.W., Rogers, S.K., Kabrisky, M., Oxley, M.E., Suter, B.W.: The multilayer perceptron as an approximation to a bayes optimal discriminant function. IEEE Trans. Neural Netw. **1**(4), 296–298 (1990)
75. Saeys, Y., Inza, I., Larrañaga, P.: A review of feature selection techniques in bioinformatics. Bioinformatics **23**(19), 2507–2517 (2007)
76. Saha, S., Bandyopadhyay, S.: A fuzzy genetic clustering technique using a new symmetry based distance for automatic evolution of clusters. In: Proceedings of the International Conference on Computing: Theory and Applications, pp. 309–314 (2007)
77. Sánchez, L., Suárez, M.R., Villar, J.R., Couso, I.: Mutual information-based feature selection and partition design in fuzzy rule-based classifiers from vague data. Int. J. Approximate Reasoning **49**, 607–622 (2008)
78. Suárez, M.R., Villar, J.R., Grande, J.: A feature selection method using a fuzzy mutual information measure. Int. J. Reasoning-Based Intell. Syst. **2**, 133–141 (2010)
79. Toivonen, H.: Sampling large databases for association rules. In: Proceedings of the 22nd Conference on Very Large Databases, vol. 96, pp. 134–145 (1996)
80. Torra, V.: Fuzzy c-means for fuzzy hierarchical clustering. In: Proceedings of the 14th IEEE International Conference on Fuzzy Systems, pp. 646–651 (2005)
81. Tsai, C.J., Lee, C.I., Yang, W.P.: A discretization algorithm based on class-attribute contingency coefficient. Inf. Sci. **178**, 714–731 (2008)
82. Verdegay, J.L., Yager, R.R., Bonissone, P.: On heuristic as a fundamental constituent of soft computing. Fuzzy Sets Syst. **159**(7), 846–855 (2008)
83. Vieira, S.M., Sousa, J.M.C., Kaymak, U.: Fuzzy criteria for feature selection. Fuzzy Set Syst. **189**, 1–18 (2012)
84. Wang, Y., Li, C., Zuo, Y.: A selection model for optimal fuzzy clustering algorithm and number of clusters based on competitive comprehensive fuzzy evaluation. IEEE Trans. Fuzzy Syst. **17**(3), 568–577 (2009)
85. Witten, I.H., Frank, E., Hall, M.: Data Mining—Practical Machine Learning Tools and Techniques. Morgan Kaufmann (2011)
86. Wong, M.L., Leung, K.S.: Data Mining Using Grammar Based Genetic Programming and Applications. Kluwer Academic Publishers (2002)
87. Yan-Qing, Y., Ju-Sheng, M., Zhou-Jun, L.: Attribute reduction based on generalized fuzzy evidence theory in fuzzy decision systems. Fuzzy Sets Syst. **170**, 64–75 (2011)
88. Yang, Y., Jia, Z., Chang, C., Qin, X., Li, T., Wang, H., Zhao, J.: An efficient fuzzy kohonen clustering network algorithm. Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery **1**, 510–513 (2008)
89. Younes, Z., Abdallah, F., Denœux, T.: Fuzzy multi-label learning under veristic variables. In: Proceedings of the IEEE International Conference on Fuzzy Systems, pp. 1–8 (2010)
90. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
91. Zadeh, L.A.: Soft computing and fuzzy logic. IEEE Softw. **11**(6), 48–56 (1994)

Soft Computing Methods in Transport and Logistics

Julio Brito, Dagoberto Castellanos-Nieves, Airam Expósito and José. A. Moreno

Abstract The current economic context generates in supply chain management greater demands for flexibility and dynamism. In addition, there is an increase in uncertainty that adds more complexity to the problems associated with planning and management. Soft Computing offers a set of methodologies capable of responding to these challenges. This work provides an overview of transport and logistics problems, as well as the most representative combinatorial optimization models. Specifically, it focuses on the treatment of uncertainty through fuzzy optimization and metaheuristics methodologies. Promising results from the use of this approach suggest emerging areas of application, which are presented and described.

1 Introduction

Logistical, transport and distribution planning have adapted to the evolution of new business organization models. Distribution strategies and transportation decisions are some of the main subjects in supply chain management and play an important role in its success because they improve service quality, reduce costs and optimize resources [20]. Supply chain management (SCM) involves all activities related to integration, planning and control of product and information flows that are generated between suppliers and clients. The supply chain can be broken down into three

J. Brito (✉) · D. Castellanos-Nieves · A. Expósito · J.A. Moreno
Grupo de Computación Inteligente, Instituto Universitario de Desarrollo Regional, Universidad de La Laguna, La Laguna, Spain
e-mail: jbrito@ull.edu.es

D. Castellanos-Nieves
e-mail: dcastell@ull.edu.es

A. Expósito
e-mail: aexposim@ull.edu.es

J.A. Moreno
e-mail: jamoreno@ull.edu.es

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_3

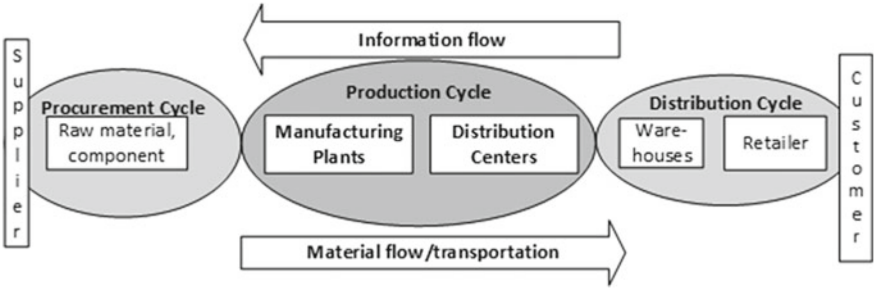


Fig. 1 Cycles of the supply chain procurement, production and distribution

cycles, procurement, production and distribution (Fig. 1). The distribution cycle refers to the activities and processes associated with storage and distribution [21].

One of the basic activities in the integral management of the supply chain is implied planning and decision making. Three levels of planning are apparent: strategic, tactical and operational [33]. Each one of these three levels is associated with different time horizons and creates a distinct set of important problems which usually correspond to optimization problems. Thus long term strategic problems include decisions on the number, size, location and capacity of the storage units and transport; tactical decisions in the medium term are the design of the distribution network, location and assignment of distribution zones/areas and the supply rate; and finally, operational aspects, studied in the short term, such as the establishment of pickup and delivery routes, and/or the organization of vehicle loading/unloading [67].

Methods/methodologies found in Soft Computing offer a useful alternative to solve problems with this type of complexity [6, 58]. The design of Intelligent Systems to aid in decision making in real settings, such as transport and logistics planning, needs to take advantage of Soft Computing methodologies [37]. The design of Intelligent Systems to aid in decision making in real settings, such as transport and logistics planning, needs to take advantage of Soft Computing methodologies. The quality of information is the most common scenario, especially in real-world applications, and this incomplete or imprecise information is reflected in the parameters and variables. Fuzzy set theory offers an appropriate methodological framework to approach this class of uncertainty, which is not the product of absence of information, nor of a random nature, but instead of the imprecise nature of the expression. Some solution techniques employ exact methods but in real-world problems, these methods do not guarantee that an optimal solution will be found. Heuristic and metaheuristics techniques are important tools constituent of Soft Computing [81] to tackle complex optimization problems. They are capable of evaluating possible alternatives and determine the preferred solution in efficient time, by means of strategies that integrate problem knowledge.

Our discussion is centered on problems and application of the levels of tactical and operational transportation planning. Transportation planning concerns the short-term planning of the distribution operations and mostly deals with the planning of deliveries to different customers. Typical considerations at this decision level are the details of delivery routes: that is, at what exact times, by which vehicle, and in what sequence customers will get their products delivered. In addition, location decision problems may have to be made on these levels.

The aim of this chapter is to analyze some relevant and emerging problems and application in transport and distribution. The application of fuzzy optimization in this field can be significant. In addition to providing an overview of transport and distribution problems and their models, the purpose is to give an overview of the fuzzy optimization and metaheuristic approach for the treatment of uncertainty in these models, to review their use and to propose new areas of application in real practical problems.

The remainder of this chapter is organized as follows. Section 2 introduces the Soft Computing based approach. Section 3 then presents a review of some problems that we have considered emerging, which are of interest for the application of Soft Computing methodologies, fuzzy optimization and metaheuristics. The chapter ends with some conclusions in Sect. 4.

2 Soft Computing Based Solution Approach

This section describes the use of an approximation that integrates specific techniques from Soft Computing, such as fuzzy optimization and metaheuristics. An outline of how these techniques are applied in the resolution of classic transport and logistics problems is also given.

2.1 Fuzzy Optimization

Fuzzy sets and systems are used to build computing systems to solve decision and optimization problems whose modeling is difficult to define accurately, managing the uncertainty and the imprecision of the available information, as well as of the formulation of preferences, restrictions and objectives expressed by decision makers. If there is imprecision in some of the formulation components of the optimization problem and we can express it with fuzzy terms then we are faced with fuzzy optimization problem. Discussions concerning solutions do not focus on their feasibility, nor if they are optimal solutions. We, in turn, have chosen to discuss the degree of feasibility and optimality of the solution. Bellman and Zadeh [10] are the authors who introduced the fundamentals for fuzzy optimization problems, where objectives and constraints can be defined in an imprecise way and characterized using membership function such as fuzzy sets. This approximation requires that the formulation and problem solutions be dealt with adequately by making use of fuzzy number representations and their operations.

An optimization problem is described as the search for the value of specific decision variables x so that identified objective functions $f(x)$ attain their optimum values. The value of the variables is subject to stated constraints $g(x)$. In these problems the objective functions are defined on a set of solutions that we will denote by X. Thus, an optimization problem can be represented by: $\max\{f(x) : g(x) \geq 0\}, \ \forall x \in X$. When some of its components are considered fuzzy, we are facing a problem of fuzzy optimization. Among all optimization problems, Linear Programming are those whose objective function and constraints are linear. The general model of Linear Programming is formulated as $\max\{f(x_j, c_j) : g(x_j, a_{ij}) \geq b_i, \ i = 1, ..., m, j = 1, ..., n, x_j \geq 0\}$. In the formulation $a_{ij} \in \mathfrak{R}^{m\times n}$ is the technological matrix, $b_i \in \mathfrak{R}^m$ the resources, $c_j \in \mathfrak{R}^n$ the costs and $x_j \in \mathfrak{R}^n$ the variables. Fuzzy Linear Programming (FLP) constitutes the basis for solving fuzzy optimization problems and their solution methods have been the subject of many studies in the fuzzy context. Different FLP models can be considered according to the fuzzy components [80]. These models can be solved in a direct and simple way, obtaining solutions that are coherent with their fuzzy nature.

(a) *Models with fuzzy constraints*. In this case there is a certain tolerance in the fulfilment constraints and consequently the feasible region can be defined as a fuzzy set. This is $\max\{f(x_j, c_j) : g(x_j, a_{ij}) \geq_f b_i, \ i = 1, ..., m, j = 1, ..., n, x_j \geq 0\}$. In particular, Verdegay [79], using the representation theorem for fuzzy sets, proves a solution which can be obtained from the auxiliary model: $\max\{f(x_j, c_j) : g(x_j, a_{ij}) \geq_f b_i + \tau_i(1 - \alpha), \ i = 1, ..., m, j = 1, ..., n, x_j \geq 0\}$, where $\alpha \in [0, 1]$ and $\tau = (\tau_1, ..., \tau_m)$ is referred to as a violation tolerance level.

(b) *Models with fuzzy cost*. In this case the model is represented by $\max\{f(x_j, c_j^f) : g(x_j, a_{ij}) \geq b_i, \ i = 1, ..., m, j = 1, ...n, x_j \geq 0\}$, where c_j^f is a fuzzy number described by its corresponding membership function $\mu_j(x)$. [23] prove that the solution can be obtained with the multi-objective auxiliary model. We can also used a simple method considering fuzzy solutions that are solved with the application of an ordered function h for the constraints [35], i.e. $\max\{f(x_j, h(c_j)) : g(x_j, a_{ij}) \geq b_i, \ i = 1, ..., m, j = 1, ...n, x_j \geq 0\}$.

(c) *Models with fuzzy coefficients in constraints*. This model considers a problem of the type $\max\{f(x_j, c_j) : g(x_j, a_{ij}^f) \geq_f b_i^f, \ i = 1, ..., m, j = 1, ...n, x_j \geq 0\}$ where the values of the technological matrix and the coefficients are fuzzy numbers and are described by its corresponding membership function. Delgado et al. [22] also include imprecision in the constraints. They propose considering fuzzy solutions that are solved with the application of an ordered function h for the constraints. The new formulation is expressed by the auxiliary problem: $\max\{f(x_j, c_j) : g(x_j, a_{ij}^f) \geq_h b_i^f + \tau_i^f(1 - \alpha), \ i = 1, ..., m, j = 1, ...n, x_j \geq 0\}$ where the symbol $\geq_h$ stands for a comparison relation between fuzzy numbers, $\alpha \in [0, 1]$ and τ_i^f is a tolerance of fuzzy nature set by the decision maker.

2.2 Evolutionary Heuristic and Metaheuristics

In Artificial Intelligence (AI), the qualifier *Heuristic* is usually applied to all those aspects related with the use of knowledge in the dynamic realization of tasks. Heuristics are used to refer to any intelligent technique, method or procedure of performing a task that is not the product of a rigorous formal analysis, but of expert knowledge about the task. In particular, the term heuristic is used to refer to a procedure that tries to provide solutions of a problem with a good performance, as regards the quality of the solutions and the resources used. Successful heuristic procedures have emerged in solving specific problems or performing difficult tasks. It has been tried to extract from them what was essential in their success in order to apply it to other problems or tasks, or in larger contexts. As has clearly occurred in various fields of AI, especially with expert systems, this line of research has contributed to the scientific development of the field of heuristics and to extend the application of its results. In this way, both specific computational techniques and resources have been obtained, as well as general design strategies for problem resolution heuristic procedures. These general strategies for constructing algorithms, which go beyond heuristics, and go further, are called Metaheuristics. Specific and elaborated heuristics for solving a simple problem in a narrow context have usually better performance than any algorithm based on metaheuristics. However, the metaheuristics tries to exploits other kind of advantages. They can improve while are used, are flexible and adaptable. They get good performance with low level of knowledge. Some recent reviews and survey show the relevance of the methods [7, 11, 68, 90].

2.3 Fuzzy Optimization in Transport and Logistics Problems

Vehicle routing, scheduling, locations and relations between them are the most important processes and decisions in transport and logistics. From a mathematical point of view, they can usually be modeled as a combinatorial optimization problem. The numerous applications of these problems include, among others, movement of goods, public transport, fresh and perishable food distributions, courier services, solid waste collection and caterers. Each problem has its own objectives, associated with cost, available resources or quality of service which in general is expressed in terms of time or distance. They also have their own specific constraints that may reflect vehicle capacity, fleet size, warehouse number and capacity, collection and/or delivery, points of loading/unloading, distances traveled and specific time windows. Most of the transportation and logistics models reviewed in this work are based on well-known problems.

Routing problems have been widely addressed in the literature and among them the Vehicle Routing Problem (VRP) is one of the most widely studied [75]. VRP is commonly defined as the problem of designing optimal delivery or collection routes for a vehicle fleet from one or several depots to a set of geographically scattered demand points, under an extensive variety of conditions. Some real-life examples, which are variants of the classic VRP with different constraints, are presented in [1]. Location problems (LP) deal with the optimal choice of a set of points for establishing certain facilities that take into account different criteria and constraints. Several models have been proposed to address these problems such as that of the P-median, P-center or covering [56]. A literature review of facility location models in the context of supply chain management is given in [49]. The Location routing problem (LRP) models and solves the facility location problem by taking into account simultaneous route planning which implies an integrated solution. Two recent surveys have been presented that describe models and variants investigated [25, 61].

In practical real-world problems decision makers use subjective knowledge or linguistic information when making decisions, measuring parameters, objectives and constraints and even when modeling the problem. In this context, a Soft Computing approach, specifically fuzzy optimization and metaheuristics, is useful for solving routing and location problems because they are flexible enough to deal with complex systems, provide acceptable approximate solutions and therefore add value. Our specific interest is in problems where various components are imprecise, treated as fuzzy and addressed with fuzzy optimization and metaheuristics.

In the literature we can find several approaches, models and solutions of the VRP considering some fuzzy component. The most widely discussed models are the ones that have imprecision demand, time travel and time services. The fuzzy optimization approach described in previous subsection are applied to VRP and some of its variants [13, 14, 74].

Several authors formulate and solve fuzzy LP, on networks [51], on networks using p-center [53], using p-median [15] or with covering [34]. Models of the fuzzy LRP problem have also been presented such as the capacitated model with fuzzy demands in [48] or a realistic version with more fuzzy components [31].

3 Applications and Emergent Problems

This section analyzes some relevant and emergent problems and application in transport and logistics, in which the application of fuzzy optimization can be significant. In addition to providing a synthesis of the existing literature on the subject, we give an overview of the problem in the context of the field to help researchers better understand the practical motivations where to apply methodologies. Thus it identifies the challenges and the direction in which future researches could be conducted in this field. The relevant and emergent problems and applications selected to realize an overview in the remainder of this section are Disaster Emergency Humanitarian Logistics, City Logistics, Green Logistics, and Tourist Trip Logistics.

3.1 Disaster Emergency Humanitarian Logistics

The International Federation of Red Cross and Red Crescent Societies defines a disaster as a sudden, calamitous event that seriously disrupts the functioning of a community or society and causes human, material, and economic or environmental losses that exceed the community's or society's ability to cope using its own resources. Though often caused by nature, disasters can have human origins. Disaster impacts more than 200 million people and produce around 75,000 fatalities every year [87].

The authors in [64] define emergency logistics as a process of planning, managing and controlling the efficient flows of relief, information, and services from the points of origin to the points of destination to meet the urgent needs of the affected people under emergency conditions. In this context, there are many challenges in logistics that differ from those encountered in commercial supply chains. Some of these challenges that introduce complexity and uncertainty are listed by [9] as:

- High level of uncertainty of demands in terms of capacities, travel times, locations, etc.
- Limited resources in large scale disasters, lack of supplies, people, technology, transportation capacity and money.
- Sudden changes in terms of demand, large resource demands in short time frames.
- Difficulties in achieving efficient and timely delivery.

In [17], an analysis of the literature is presented, showing the main optimization models used in emergency logistics. Facility location, location-evacuation, location with relief distribution and pre-positioning, relief distribution and casualty transportation, resource allocation, commodity flow, resource allocation and commodity flow, and other models are others just some of the cited models.

A survey of recent advances in bio-inspired meta-heuristics, including genetic algorithms, particle swarm optimization, ant colony optimization, etc., for solving emergency transportation problems is also presented in [92].

Optimization problems of emergency logistic involve complex systems introducing fuzzy sets and systems. In the literature there are several papers related to fuzzy location and routing problem, for instance, in [66], a hybrid fuzzy optimization methodology to solve the large-scale disaster relief distribution problem is presented. The solution approach is made up of three steps. In the first step fuzzy clustering techniques are used to classify the damaged areas, while the second and third steps use FLP to deal with lack of resources. The authors in [65] provide a hybrid fuzzy clustering optimization approach to the operation of an emergency logistics co-distribution center responding to the urgent relief demands in the crucial rescue period.

The fuzzy LRP is a research area with several papers in emergency logistics. In [84] a model considering fuzzy demands of relief materials, timeliness and limited resources is proposed. The objective function of the model minimizes the total cost and the relief time of the system.

Other papers focus on emergency logistics transportation path optimization using fuzzy approaches. In [89] a multi-objective optimization model for emergency logistics transportation path is proposed. Factors such as transportation time, safety factor and transportation costs are described using trapezoidal fuzzy numbers in order to deal with the uncertainty of these factors.

3.2 City Logistics

Urban mobility plays a key role in the promotion of the sustainable urban development of a city. In particular, an efficient freight transport system is required as it plays a significant role in its competitiveness and represents an important element for the local economy regarding the employment and income that it generates [63]. Logistics activities and operations, especially transportation and distribution of goods, now receive a specific treatment which is known as urban logistics or city logistics. This concept has its origins in the 1980s, it has become in the last decades more relevant by the development of cities, the growing demands for supply that is more efficient and concern about the negative impacts of it. Taniguchi et al. [73] defined city logistics as the process for totally optimizing the logistics and transport activities by private companies with support of advanced information systems in urban areas considering the traffic environment, the traffic congestion, the traffic safety and the energy savings within the framework of a market economy [73].

Urban distribution and transport of goods is an important part of urban logistics. In cities complicated problems arise related to urban freight transport: demand of higher levels of service in terms of time, need for better services with lower costs for customers, use of fewer vehicles, better utilization of vehicle capacity, reduction of negative environmental impacts, lower energy consumption, reduction of noise, contribution to traffic congestion, the use of alternative energies and improving safety [26]. In the next coming years, changes from increased e-commerce and home delivery will be apparent. This growth reinforces the general trend in logistics towards smaller consignments, single orders and thus higher delivery frequency and an increase in vehicle movements within cities [83].

In urban areas, goods distribution services are the most important transportation and logistics activities and are usually called the last mile. Clear examples of these are courier express and parcel services food delivery, perishable products, milk and newspaper, urban solid waste collection and emergency transport. Surveys about these problems are available in [16, 41].

Important areas in urban freight routes are concerned with reduction of fuel consumption and emissions [70], and the use of night delivery schemes [18]. In contemporary living the travel speed between locations varies throughout the day according to traffic conditions, especially in urban areas. Therefore, it is necessary to adapt the models with time-dependence for routing planning [27]. The effects of e-commerce on the urban freight transport using vehicle routing and scheduling problem model are studied in [36, 72].

The design and development of city logistics systems requires the availability of the models for the location of the logistics centers. The location of warehouse, distribution centers and/or consolidation centers should be appropriately determined for optimal operations [54]. On the other hand, depot location and vehicle routing are two interdependent decisions and there are many authors who consider the integrated problem of locating distribution centers in urban areas and the corresponding freight distribution [52]. Recently, the authors in [38] investigate the combined impact of depot location, fleet composition and routing decisions on vehicle emissions in city logistics. When real-time information is available, dynamic models can support the developed systems. For example, in [29] on-line re-optimization based on current traffic information and soft time constraints are proposed.

All of these problems are subject to the uncertainty of the environment. In some situations logistic problems cannot exactly specify attributes as either deterministic numbers or probabilistic random variables and it is natural and realistic to express vagueness and ambiguity in fuzzy terms and to solve the problem with a fuzzy approach [88]. In the literature, Soft Computing methodologies have been considered in routing, scheduling and location. But integrating such methods into city logistics has not been fully considered and needs to be addressed. Only a few references are found in the literature. In [45] a postal delivery in agglomerations with a large number of customers modelled as a street routing problem is solved using fuzzy clustering. Customer clustering is also used for VRP and is solved in the framework of urban freight transport [85]. Another paper studies the dynamic operational environment of courier service with fuzzy time windows [44]. Location models [86] and specifically the design of e-commerce distribution systems [40] complete the papers found in the literature.

3.3 Green Logistics

For decades, the main goal of logistics has been to improve its objectives from a purely economic point of view. Through this period logistics professionals have perpetuated this economic and commercial paradigm by allowing organization and management logistics to focus on maximizing economic profitability. Economic performance did not include costs such as environmental and social impact, instead it only considered operational economic costs.

Nowadays there is a growing demand for multi-objective metrics associated with logistical processes, where reduced operational costs and the negative impact of the environment are some of the most common. As regards the last objective, logistic stakeholders are pressured from public administrations and government to reduce the environmental impact of their logistics operations. The environmental impact of logistics companies can be measured in different ways such as generation of noise and vibrations, air quality and the contribution to global warming.

Mckinnon et al. [47] present a classification of the main pillars that compose what we call green logistics:

- Reducing Freight transport externalities. The largest part of the research on the environmental impact of logistics is due to growth of freight traffic volumes. As a consequence, the search for ways of reducing freight-related externalities is a priority.
- City Logistics. The study of transport efficiency and environmental impact of logistics in urban areas. This topic will be further developed in later sections.
- Reverse Logistics. The logistic associated with waste product and packaging for reuse, is a promising area of green logistics. The logistics of waste management help to increase the proportion of waste material that is recycled and reused.
- Corporate environmental strategies. Companies have adopted and improved strategies to reduce environmental impact and manage a balanced relationship with nature. These strategies generate a wide range of actions on logistics operations of the company.
- Green Supply Chain Management. This pillar can be defined as the alignment and integration of environmental management within supply chain management. This is based on that part of the environmental impact that is extended beyond their own structure.

It is possible to find several surveys and literature reviews related to green logistics. In [93] a review of combinatorial optimization problems related with green logistics and meta-heuristics in the swarm intelligence field is proposed. A review of Green Logistic Vehicle Routing Problems (GVRP) and a discussion of the next wave of research into GVRP is presented in [43].

In general, the complex infrastructure of logistics and working in a dynamic business environment imposes a high degree of uncertainty in the logistics process that affects its overall performance. More specifically, there is a complex system present in green logistics with highly imprecise parameters and environmental factors with a complexity that requires fuzzy sets to be represented.

In the literature, it has been proven that green logistics belong to a wide range of practical application areas. In most cases, these optimization problems involve complex systems introducing fuzzy sets and systems. A literature review and a discussion of the applications of fuzzy green logistics focus on fuzzy optimization in green transport are provided in the remaining part of this section.

There are several papers in literature related with green logistics and fuzzy approaches, specifically in green routing problems. A transport spatial decision support model for the optimization of green routes for city logistics centers is presented in [57], where the solution integrates a multi-criteria method of Weighted Linear Combination and the modified Dijkstra algorithm with a geographic information system for processing spatial data in order to minimize the environmental impact of the routes. In freight transportation activities, [28] puts forward the use of green logistics in order to reduce the negative impact on the environment considering demand and travel time uncertainty.

Other papers focus on aspects such as reverse logistics and green scheduling. In [62], a bi-objective mathematical model in the distribution of perishable products with specified expiration date in inventory routing problem is presented. In order to solve the model the Torabi-Hassini method based on a fuzzy approach is applied. A green train scheduling model with a fuzzy multi-objective optimization algorithm is presented in [42]. In order to solve the problem and obtain non-dominated solutions which has an equally satisfactory degree on both objectives, a fuzzy mathematical programming approach is proposed.

3.4 Tourist Trip Logistics

The Tourist Trip Design Problem, (TTDP) [78] arises when a tourist visiting a tourist area, for one or more days, and is interested in visiting a number of points of interest (POI). The problem more specifically deals with the choice of the POIÂt's of the trip and the order to follow each day. Each one of these POIs is associated with several features that are taken into account in the selection. The main features are the minimum time required for the visit, the days and hours of operation, the cost of the activities within the visit and some indicators of profit or degree of satisfaction that could be perceived by the tourist for the visit. Information regarding distance, travel time and cost between the POIs, and between these points and the hotel or stay residence of the tourist (start and end point of the trips), must also be taken into account. This information together with some information about the tourist, such as his/her preferences, budget and time limitations must be used to decide the trip selected for each day of the stay at destination. The corresponding optimization problems have received increasing interest in the tourist management and service in order to be incorporated to recommenders, tourist planning tools and electronic guides. Since most of the features that have to be used are subjective or subject to some level of imprecision and vagueness, the fuzzy techniques and points of view have been used.

The problems may be complicated and made more realistic by considering additional features and constraints. Some of them are maximum budget (daily or complete stay at destination), specific requirements on the minimum and/or maximum number of days that the tourist visits to the POIs within a certain category (restaurants, beaches, historic sites, nature facilities, etc.), or on the number of visits to POIs of a category some days. Travel times that depend on traffic congestion, weather conditions, or the time of the day when traveling. Other realistic variants arise when some of the POIs have time window constraints and the time used to visit them have to be taken into account in the cost or profit of the visit. Finally, in realistic cases evaluation of the profits or interest of the POIs depend on the already visited POIs or some additional information [5].

The literature firstly distinguishes between problems with only one tour and with several tours. Most research considers two opposed features or criteria: the profit and the cost. In the single tour problem, the main variants are the Profitable Tour Problem (PTP) where the objective is to maximize the difference between the profit and the total cost that has been considered in [24], the Prize Collecting Travelling Salesman Problem (PCTSP) where the objective is to minimize the travel cost subject to a minimum profit [2], and the Orienteering Problem (OP) where the objective is to maximize the total profit subject to a limit on the travel costs that was introduced by [77].

Problems with multiple tours are usually known as Vehicle Routing Problems with Profit (VRPP), and the simplest version is the well-known Team Orienteering Problem (TOP), which is an extension of the OP in which there is a fixed number m of routes and m also corresponds to the number of days available to tourists.

There are other problems using different objective functions, such as the Prize-Collecting VRP (PCVRP) [71] and the Capacitated Profitable Tour Problem (CPTP) [3]. In PCVRP, the objective function is given by the combination of three different objectives: minimizing the travel distance, minimizing the number of vehicles used and maximizing profits collected, while in CPTP the goal is to minimize the difference between the total harvest profit and the total trip cost.

The Team Orienteering Problem has been extensively studied in the literature [4] and several versions, such as the Team Orienteering Problem with Time Windows (TOPTW) [76] and Time Dependent Team Orienteering Problem with Time Windows [30].

The use of metaheuristics and fuzzy approaches are quite common in wide range of recent TTDP and other routing problems. The early work [46] considers a fuzzy routing problem for sightseeing. A Genetic Algorithm for the VRPTW with fuzzy demand is applied in [91]. In [55] a supplier selection model using fuzzy logic is developed. Several multi-objective metaheuristics are used in [69] to solve VRP with fuzzy demands. Several authors [19, 59, 60] apply a Particle Swarm Optimization for a VRP with fuzzy demands and [8] use Genetic Algorithm. The paper [94] deal with the CVRPTW with fuzzy travel time and demand using a hybrid between Ant Colony and Genetic Algorithm. A fuzzy capacitated location routing problem is solved in [32] by applying a Simulating Annealing with a mutation operator. Recently [82] apply fuzzy optimization to the orienteering problem, [39] apply a fuzzy Ant Colony system to solve the dynamic vehicle routing problem with uncertain service time, and [50] uses fuzzy number comparisons to deal with VRPTW with fuzzy scores. Finally, [12] apply a GRASP for solving the TOP with fuzzy scores and constraints.

4 Conclusions

This review has focused on two major areas of Soft Computing applied to optimization: metaheuristics and fuzzy optimization. Works in metaheuristics are numerous and often applied in emerging logistics areas. Relevant works with metaheuristics outperform other methods, such as exact algorithms. Fuzzy optimization methods are less frequent in the literature, providing additional research opportunities for Soft Computing in a very open and promising field.

Acknowledgements This work has been partially funded by the Spanish Ministry of Economy and Competitiveness with FEDER funds (TIN2015-70226-R) and supported by Fundación Cajacanarias research funds (project 2016TUR19). Contributions from A. Expósito is supported by the ACIISI of the Gobierno de Canarias and FSE.

References

1. Anbuudayasankar, S.P., Ganesh, K., Mohapatra, S.: Models for Practical Routing Problems in Logistics. Springer International Publishing (2014)
2. Archetti, C., Bianchessi, N., Speranza, M.G.: Optimal solutions for routing problems with profits. Discrete Appl. Math. **161**(4), 547–557 (2013)
3. Archetti, C., Feillet, D., Hertz, A., Speranza, M.G.: The capacitated team orienteering and profitable tour problems. J. Oper. Res. Soc. **60**(6), 831–842 (2009)
4. Archetti, C., Hertz, A., Speranza, M.G.: Metaheuristics for the team orienteering problem. J. Heuristics **13**(1), 49–76 (2007)
5. Archetti, C., Speranza, M.G., Vigo, D.: Vehicle routing: problems, methods, and applications, vol. 18, chapter Vehicle Routing Problem with Profits, pp. 273. SIAM (2014)
6. Avineri, E.: Soft computing applications in traffic and transport systems: a review. In: Hoffmann, F., Kappen, M., Klawonn, F., Roy, R. (eds.) Soft Computing: Methodologies and Applications. Advances in Soft Computing, vol. 32, pp. 17–25. Springer, Berlin (2005)
7. Baghel, M., Agrawal, S., Silakari, S.: Survey of metaheuristic algorithms for combinatorial optimization. Int J Comput Appl **58**(19) (2012)
8. Bahri, O., Amor, N.B., El-Ghazali, T.: Optimization algorithms for multi-objective problems with fuzzy data. In: 2014 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM), pp. 194–201. IEEE (2014)
9. Balcik, B., Beamon, B.M.: Facility location in humanitarian relief. Int. J. Logistics Res. Appl. **11**(2), 101–121 (2008)
10. Bellman, R.E., Zadeh, L.A.: Decision-making in a fuzzy environment. Manage. Sci. **17**(4), B–141 (1970)
11. BoussaïD, I., Lepagnot, J., Siarry, P.: A survey on optimization metaheuristics. Inf. Sci. **237**, 82–117 (2013)
12. Brito, J., Expósito, A., Moreno, J.A.: Solving the team orienteering problem with fuzzy scores and constraints. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1614–1620. IEEE (2016)
13. Brito, J., Moreno, J.A., Verdegay, J.L.: Fuzzy optimization in vehicle routing problems. In: IFSA/EUSFLAT Conference, pp. 1547–1552 (2009)
14. Brito, J., Moreno, J.A., Verdegay, J.L.: Transport route planning models based on fuzzy approach. Iran. J. Fuzzy Syst. **9**(1), 141–158 (2012)
15. Cadenas, J., Canas, M., Garrido, M., Ivorra, C., Liern, V.: Soft-computing based heuristics for location on networks: the p-median problem. Appl. Soft Comput. **11**(2), 1540–1547 (2011). The Impact of Soft Computing for the Progress of Artificial Intelligence
16. Cattaruzza, D., Absi, N., Feillet, D., González-Feliu, J.: Vehicle routing problems for city logistics. EURO J. Transp. Logistics 1–29 (2015)
17. Caunhye, A.M., Nie, X., Pokharel, S.: Optimization models in emergency logistics: a literature review. Socio-Econ. Plann. Sci. **46**(1), 4–13 (2012). Special Issue: Disaster Planning and Logistics: Part 1
18. Cavar, I., Kavran, Z., Jolic, N.: Intelligent transportation system and night delivery schemes for city logistics. Comput. Technol. Appl. **2**(9), 782–787 (2011)
19. Chen, J.-Q., Li, W.-L., Murata, T.: Particle swarm optimization for vehicle routing problem with uncertain demand. In: 2013 4th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 857–860. IEEE (2013)

20. Cooper, M.C., Lambert, D.M., Pagh, J.D.: Supply chain management: more than a new name for logistics. Int. J. Logistics Manag. **8**(1), 1–14 (1997)
21. Croom, S., Romano, P., Giannakis, M.: Supply chain management: an analytical framework for critical literature review. Eur. J. Purchasing Supply Manag. **6**(1), 67–83 (2000)
22. Delgado, M., Verdegay, J., Vila, M.: A general model for fuzzy linear programming. Fuzzy Sets Syst. **29**, 21–29 (1989)
23. Delgado, M., Verdegay, J.L., Vila, M.A.: Imprecise costs in mathematical programming problems. Control Cybernet **16**(2), 113–121 (1987)
24. Dell'Amico, M., Maffioli, F., Värbrand, P.: On prize-collecting tours and the asymmetric travelling salesman problem. Int. Trans. Oper. Res. **2**(3), 297–308 (1995)
25. Drexl, M., Schneider, M.: A survey of variants and extensions of the location-routing problem. Eur. J. Oper. Res. **241**(2), 283–308 (2015)
26. Ehmke, J.: Integration of information and optimization models for routing in city logistics, vol. 177. Springer Science & Business Media (2012)
27. Ehmke, J.F., Steinert, A., Mattfeld, D.C.: Advanced routing for city logistics service providers based on time-dependent travel times. J. Comput. Sci. **3**(4), 193–205 (2012). City Logistics
28. Eshtehadi, R., Fathian, M., Demir, E.: Robust solutions to the pollution-routing problem with demand and travel time uncertainty. Transp. Res. Part D: Transp. Environ. **51**, 351–363 (2017)
29. Flamini, M., Nigro, M., Pacciarelli, D.: The value of real-time traffic information in urban freight distribution. J. Intell. Transp. Syst. (2017). In press
30. Gavalas, D., Konstantopoulos, C., Mastakas, K., Pantziou, G., Vathis, N.: Heuristics for the time dependent team orienteering problem: application to tourist route planning. Comput. Oper. Res. **62**, 36–50 (2015)
31. Ghaffari-Nasab, N., Ahari, S.G., Ghazanfari, M.: A hybrid simulated annealing based heuristic for solving the location-routing problem with fuzzy demands. Scientia Iranica **20**(3), 919–930 (2013)
32. Golozari, F., Jafari, A., Amiri, M.: Application of a hybrid simulated annealing-mutation operator to solve fuzzy capacitated location-routing problem. Int. J. Adv. Manuf. Technol. **67**(5–8), 1791–1807 (2013)
33. Gunasekaran, A., Kobu, B.: Performance measures and metrics in logistics and supply chain management: a review of recent literature (1995–2004) for research and applications. Int. J. Prod. Res. **45**(12), 2819–2840 (2007)
34. Guzmán, V.C., Masegosa, A.D., Pelta, D.A., Verdegay, J.L.: Fuzzy models and resolution methods for covering location problems: an annotated bibliography. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **24**(04), 561–591 (2016)
35. Herrera, F., Verdegay, J.: Three models of fuzzy integer linear programming. Eur. J. Oper. Res. **83**(3), 581–593 (1995)
36. Hong, X., Jingjing, Q., Xingli, T.: B2c e-commerce vehicle delivery model and simulation. Inf. Technol. J. **12**(20), 5891 (2013)
37. Ko, M., Tiwari, A., Mehnen, J.: A review of soft computing applications in supply chain management. Appl. Soft Comput. **10**(3), 661–674 (2010)
38. Koc, C., Bekta, T., Jabali, O., Laporte, G.: The impact of depot location, fleet composition and routing on emissions in city logistics. Transp. Res. Part B: Methodol. **84**, 81–102 (2016)
39. Kuo, R., Wibowo, B., Zulvia, F.: Application of a fuzzy ant colony system to solve the dynamic vehicle routing problem with uncertain service time. Appl. Math. Modell. **40**(23), 9990–10001 (2016)
40. Lau, H.C.W., Jiang, Z.Z., Ip, W.H., Wang, D.W.: A credibility-based fuzzy location model with hurwicz criteria for the design of distribution systems in b2c e-commerce. Comput. Ind. Eng. **59**(4), 873–886 (2010)
41. Lewczuk, K., Żak, J., Pyza, D., Jacyna-Gołda, I.: Vehicle routing in an urban area: environmental and technological determinants. WIT Trans. Built Environ. **130**, 373–384 (2013)
42. Li, X., Wang, D., Li, K., Gao, Z.: A green train scheduling model and fuzzy multi-objective optimization algorithm. Appl. Math. Modell. **37**(4), 2063–2073 (2013)

43. Lin, C., Choy, K., Ho, G., Chung, S., Lam, H.: Survey of green vehicle routing problem: past and future trends. Expert Syst. Appl. **41**(4, Part 1), 1118–1138 (2014)
44. Lin, C., Choy, K., Ho, G., Lam, H., Pang, G.K., Chin, K.: A decision support system for optimizing dynamic courier routing operations. Expert Syst. Appl. **41**(15), 6917–6933 (2014)
45. Matis, P., Koháni, M.: Very large street routing problem with mixed transportation mode. CEJOR **19**(3), 359–369 (2011)
46. Matsuda, Y., Nakamura, M., Kang, D., Miyagi, H.: A fuzzy optimal routing problem for sightseeing. IEEJ Trans. Electron. Inf. Syst. **125**, 1350–1357 (2005)
47. McKinnon, P., Cullinane, S., Whiteing, A., Browne, P.: Green Logistics: Improving the Environmental Sustainability of Logistics. Kogan Page (2010)
48. Mehrjerdi, Y.Z., Nadizadeh, A.: Using greedy clustering method to solve capacitated location-routing problem with fuzzy demands. Eur. J. Oper. Res. **229**(1), 75–84 (2013)
49. Melo, M., Nickel, S., da Gama, F.S.: Facility location and supply chain management: a review. Eur. J. Oper. Res. **196**(2), 401–412 (2009)
50. Mendez, C.E.C.: Team Orienteering Problem with Time Windows and Fuzzy Scores. PhD thesis, National Taiwan University of Science and Technology (2016)
51. Moreno, J.A., Vega, J.M., Verdegay, J.L.: Fuzzy location problems on networks. Fuzzy Sets Syst. **142**(3), 393–405 (2004)
52. Muñoz-Villamizar, A., Montoya-Torres, J.R., Juan, A.A., Cáceres-Cruz, J.: A simulation-based algorithm for the integrated location and routing problem in urban logistics. In: 2013 Winter Simulations Conference (WSC), pp. 2032–2041 (2013)
53. Nayeem, S.M.A., Pal, M.: The p-center problem on fuzzy networks and reduction of cost. Iran. J. Fuzzy Syst. **5**(1), 1–26 (2008)
54. Olsson, J., Woxenius, J.: Localisation of freight consolidation centres serving small road hauliers in a wider urban area: barriers for more efficient freight deliveries in gothenburg. J. Transp. Geogr. **34**, 25–33 (2014)
55. Ordoobadi, S.M.: Development of a supplier selection model using fuzzy logic. Supply Chain Manage. Int. J. **14**(4), 314–327 (2009)
56. Owen, S.H., Daskin, M.S.: Strategic facility location: a review. Eur. J. Oper. Res. **111**(3), 423–447 (1998)
57. Pamučar, D., Gigović, L., Ćirović, G., Regodić, M.: Transport spatial model for the definition of green routes for city logistics centers. Environ. Impact Assess. Rev. **56**, 72–87 (2016)
58. Peidro, D., Mula, J., Poler, R., Verdegay, J.-L.: Fuzzy optimization for supply chain planning under supply, demand and process uncertainties. Fuzzy Sets Syst. **160**(18), 2640–2657 (2009)
59. Peng, Y., Chen, J.: Vehicle routing problem with fuzzy demands and the particle swarm optimization solution. In: 2010 International Conference on Management and Service Science (MASS), pp. 1–4. IEEE (2010)
60. Peng, Y., Qian, Y.-M.: A particle swarm optimization to vehicle routing problem with fuzzy demands. J. Convergence Inf. Technol. **5**(6), 112–119 (2010)
61. Prodhon, C., Prins, C.: A survey of recent research on location-routing problems. Eur. J. Oper. Res. **238**(1), 1–17 (2014)
62. Rahimi, M., Baboli, A., Rekik, Y.: Sustainable inventory routing problem for perishable products by considering reverse logistic. IFAC-PapersOnLine **49**(12), 949–954 (2016)
63. Russo, F., Comi, A.: A classification of city logistics measures and connected impacts. Procedia—Soc. Behav. Sci. **2**(3), 6355–6365 (2010)
64. Sheu, J.-B.: Challenges of emergency logistics management. In: Transportation Research Part E: Logistics and Transportation Review, vol. 43, no. 6, pp. 655–659 (2007). Challenges of Emergency Logistics Management
65. Sheu, J.-B.: An emergency logistics distribution approach for quick response to urgent relief demand in disasters. In: Transportation Research Part E: Logistics and Transportation Review, vol. 43, no. 6, pp. 687–709 (2007). Challenges of Emergency Logistics Management
66. Sheu, J.-B., Chen, Y.-H., Lan, L.W., et al.: A novel model for quick response to disaster relief distribution. Proc. Eastern Asia Soc. Transp. Stud. **5**, 2454–2462 (2005)

67. Simchi-Levi, D., Chen, X., Bramel, J.: The logic of logistics. Algorithms, and Applications for Logistics and Supply Chain Management, Theory (2005)
68. Sörensen, K.: Metaheuristics-the metaphor exposed. Int. Trans. Oper. Res. **22**(1), 3–18 (2015)
69. Sulieman, D., Jourdan, L., Talbi, E.-G.: Using multiobjective metaheuristics to solve vrp with uncertain demands. In: 2010 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8. IEEE (2010)
70. Suzuki, Y.: A new truck-routing approach for reducing fuel consumption and pollutants emission. Transp. Res. Part D: Transp. Environ. **16**(1), 73–77 (2011)
71. Tang, L., Wang, X.: Iterated local search algorithm based on very large-scale neighborhood for prize-collecting vehicle routing problem. Int. J. Adv. Manuf. Technol. **29**(11), 1246–1258 (2006)
72. Taniguchi, E., Kakimoto, Y.: Modelling effects of e-commerce on urban freight transport, chapter Chapter 10, pp. 135–146. emeraldinsight (2004)
73. Taniguchi, E., Thompson, R., Yamada, T., Duin, R.V.: City Logistics: Network Modelling and Intelligent Transport Systems. Pergamon (2001)
74. Torres, I., Cruz, C., Verdegay, J.L.: Solving the truck and trailer routing problem with fuzzy constraints. Int. J. Comput. Intell. Syst. **8**(4), 713–724 (2015)
75. Toth, P., Vigo, D.: Vehicle Routing. Society for Industrial and Applied Mathematics. Philadelphia, PA (2014)
76. Tricoire, F., Romauch, M., Doerner, K.F., Hartl, R.F.: Heuristics for the multi-period orienteering problem with multiple time windows. Comput. Oper. Res. **37**(2), 351–367 (2010)
77. Tsiligirides, T.: Heuristic methods applied to orienteering. J. Oper. Res. Soc. 797–809 (1984)
78. Vansteenwegen, P., Oudheusden, D.V.: The mobile tourist guide: an or opportunity. OR Insight **20**(3), 21–27 (2007)
79. Verdegay, J.: Fuzzy Information and Decision Processes, chapter Fuzzy mathematical programming. North-Holland (1982)
80. Verdegay, J.L.: Fuzzy optimization: models, methods and perspectives. In: In proceeding 6th IFSA-95 World Congress, pp. 39–71 (1995)
81. Verdegay, J.L., Yager, R.R., Bonissone, P.P.: On heuristics as a fundamental constituent of soft computing. Fuzzy Sets Syst. **159**, 846–855 (2008)
82. Verma, M., Shukla, K.K.: Application of fuzzy optimization to the orienteering problem. Adv. Fuzzy Syst. **2015**, 8 (2015)
83. Visser, J., Nemoto, T., Browne, M.: Home delivery and the impacts on urban freight transport: a review. Procedia—Soc. Behav. Sci. **125**, 15–27 (2014)
84. Wang, S., Ma, Z., Zhuang, B.: Fuzzy location-routing problem for emergency logistics systems. Comput. Modell. New Technol. **18**(2), 265–273 (2014)
85. Wang, Y., Ma, X., Xu, M., Wang, Y., Liu, Y.: Vehicle routing problem based on a fuzzy customer clustering approach for logistics network optimization. J. Intell. Fuzzy Syst. **29**, 1427–1442 (2015)
86. Wang, Y., Ma, X.L., Wang, Y.H., Mao, H.J., Zhang, Y.: Location optimization of multiple distribution centers under fuzzy environment. J. Zhejiang Univ. Sci. A **13**(10), 782–798 (2012)
87. Wassenhove, L.N.V.: Humanitarian aid logistics: supply chain management in high gear. J. Oper. Res. Soc. (2006)
88. Wong, B.K., Lai, V.S.: A survey of the application of fuzzy set theory in production and operations management: 1998–2009. Int. J. Prod. Econ. **129**(1), 157–168 (2011)
89. Xiao, S.C., Wu, J.F., He, H., Yang, Z.D., Shen, X.: An emergency logistics transportation path optimization model by using trapezoidal fuzzy. In: 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 199–203 (2014)
90. Xiong, N., Molina, D., Ortiz, M.L., Herrera, F.: A walk into metaheuristics for engineering optimization: principles, methods and recent trends. Int. J. Comput. Intell. Syst. **8**(4), 606–636 (2015)
91. Xu, J., Goncalves, G., Hsu, T.: Genetic algorithm for the vehicle routing problem with time windows and fuzzy demand. In: Evolutionary Computation, 2008, pp. 4125–4129. IEEE (2008)

92. Zhang, M.-X., Zhang, B., Zheng, Y.-J.: Bio-inspired meta-heuristics for emergency transportation problems. Algorithms **7**(1), 15–31 (2014)
93. Zhang, S., Lee, C., Chan, H., Choy, K., Wu, Z.: Swarm intelligence applied in green logistics: a literature review. Eng. Appl. Artif. Intell. **37**, 154–169 (2015)
94. Zulvia, F.E., Kuo, R., Hu, T.-L.: Solving cvrp with time window, fuzzy travel time and demand via a hybrid ant colony optimization and genetic algortihm. In: 2012 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8. IEEE (2012)

Applications of Soft Computing in Intelligent Transportation Systems

Antonio D. Masegosa, Enrique Onieva, Pedro Lopez-Garcia and Eneko Osaba

Abstract Intelligent Transportation Systems emerged to meet the increasing demand for more efficient, reliable and safer transportation systems. These systems combine electronic, communication and information technologies with traffic engineering to respond to the former challenges. The benefits of Intelligent Transportation Systems have been extensively proved in many different facets of transport and Soft Computing has played a major role in achieving these successful results. This book chapter aims at gathering and discussing some of the most relevant and recent advances of the application of Soft Computing in four important areas of Intelligent Transportation Systems as autonomous driving, traffic state prediction, vehicle route planning and vehicular ad hoc networks.

1 Introduction

New trends in business, commerce or leisure have increased the demand for more efficient, reliable and safer transportation systems. This fact is claimed by different national and international institutions, such as the OECD [49] or the European Commission [20], just to name but a few. Some of the reasons behind the increasing

A.D. Masegosa (✉) · E. Onieva · P. Lopez-Garcia · E. Osaba
Faculty of Engineering, University of Deusto, 48007 Bilbao, Spain
e-mail: ad.masegosa@deusto.es

E. Onieva
e-mail: enrique.onieva@deusto.es

P. Lopez-Garcia
e-mail: pedro.lpzgrc@deusto.es

E. Osaba
e-mail: e.osaba.es@deusto.es

A.D. Masegosa · E. Onieva · P. Lopez-Garcia · E. Osaba
DeustoTech-Fundacion Deusto, Deusto Foundation, 48007 Bilbao, Spain

A.D. Masegosa
IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_4

importance of improving transportation systems are the offshore outsourcing of production, the adoption of just-in-time distribution systems, the tight scheduling of personal and freight activities, the broadening of international trade, the large number of people living in cities or the big amount of Green House Gases emission caused by transport.

The development of Intelligent Transportation Systems (ITS) [18] is one of the major areas of research that work in addressing these issues nowadays. ITS is a discipline that combines telecommunication, electronics, information technologies and traffic engineering methodologies to provide innovative services associated to different modes of transport as well as traffic management in order to offer the users more information and safety and to allow a more efficient and effective use of transport networks. The benefits of these systems have been successfully proved in many different transport environments [33].

Since their beginning around 1930 with the first electric traffic signals, ITS have coped virtually all facets of transport given rise to different types of systems to respond to the different problems that appear in each of those facets. Some of these types of ITS are [23]:

- *Advanced Traffic Management Systems (ATMSs)* that aim at improving traffic service quality by collecting data, supporting decision making to operators and control traffic in real-time with different systems.
- *Advanced Travelers Information Systems (ATISs)* which are design to help travelers in the different stages of their trips by providing them information in real-time about the best route to their destiny, the most appropriate schedule or transport media, etc.
- *Commercial Vehicles Operation Systems* whose objective is to increase safety and efficiency in commercial vehicles fleets, by combining different ITS technologies with the intention of improving the management and control of vehicles as well as the information available for drivers and decision makers.
- *Advanced Public Transportations Systems (APTSs)* that pursue the enhancement of the operation of public transportation media (subway, tram, bus, etc.) with the joint use of technologies from ATISs and ATMSs.
- *Advanced Vehicles Control Systems (AVCSs)* aims at assisting, alerting and taking the whole or part of vehicle driving with the aid of several in-vehicle sensors, computers and/or communication networks.

Soft Computing (SC) [75] has played a major role in the success of ITS in recent years, especially because of the bigger amount of data provided and collected from several sources by the different stakeholders involved in these systems as governments, industry and citizens [92]. ITS are an environment very appropriated to applied SC techniques because the information handle present most of the features for which SC was designed for. For example, the sensors usually present imprecision in their measures; traffic is strongly affected by factors with a high uncertainty as weather; and the decision making should take into account drivers' or other users' preferences which subject to a high vagueness and subjectivity. For these reasons, techniques such as fuzzy sets, neural networks, metaheuristics or probabilistic reasoning have been widely used by the research community in ITS.

This book chapter aims at gathering some of the most relevant and recent advances of the application of SC in ITS in order to serve as a guide for student, researchers and practitioners interested in this field. Concretely, we focus on four important areas of ITS as the AVCSs for autonomous driving, the ATISs for the prediction of the future state of the traffic, the CVOSs for the planning of routes of fleets of vehicles, as well as one of the most important key enable technologies of future ITS, vehicular ad hoc networks.

The manuscript is organized as follows. The next four sections are devoted to review the application of SC in each of the ITS areas aforementioned in order of appearance. After that, in the last section of the chapter, we discuss the main conclusions drawn from the works reviewed.

2 Soft Computing in Autonomous Driving

Autonomous driving has been one of the most benefited fields in ITS of the application of SC, since, until recently, autonomous driving remained like one of those problems that humans were able to manage better than machines [13].

The European Union has an ambitious road safety target for this decade: halving the number of road deaths between 2010 and 2020. In 2014, almost 25,700 road fatalities were reported in the EU, this is around 1% fewer deaths than reported in 2013 and 18% fewer than in 2010 [1].

Automation, and in particular digitalization of driving will change road transport in a way which will be viewed as a revolution in the field of mobility. As human error is the main reason for road traffic accidents, controlling the driving by a computer is expected to make future road transport safer and more secure. A fully automated mobility of vehicles in roads will have incredible potential impact in the society as known until now. Benefits from such total automation of the vehicles will derive in evident profits for society lowering costs and increasing safety, but also will provide deep changes in the ways people and goods move around cities.

Car sharing [8] and car pooling [82] are two examples of emerging paradigms of mobility that can deeply impact the mobility, if come accompanied by autonomous driving of vehicles involved in the business models. Some authors study social tendency of population to owning a vehicle in property, and question about the effect that autonomous vehicles will cause in social perception of owning a car [76]. Philosophical and ethical sciences are also influenced by the emergence of autonomous vehicle applications, and researchers try to answer if humans are ready for utilitarian autonomous vehicles [11]. As can be seen, not only technological but also societal and business, among other fields of research, are involved in the study of the automation of vehicles.

In one hand, both the upgrade and lowering of the sensor and equipment necessary for the processing of information needed by an autonomous vehicle to take a decision have made this field to receive interest from research groups from the entire world. In the other, methods and algorithms able to deal in each one of the stages involved in the driving of a vehicle (e.g. signal processing, decision making, control, communication or planning) are performing really well nowadays. Both equipment and methods are making nearer the day in which a fully autonomous driving is available for society.

It is known that the techniques under the topic of Soft Computing have a strong capability of learning and cognition, as well as a good tolerance to uncertainty and imprecision. Due to these properties they can be applied successfully to problems derived by the driving of a vehicle. Methods associated to the field of Soft Computing have been naturally used by researchers to take steps through the development of autonomous vehicles.

In the remainder part of this section, some of the most relevant applications of Soft Computing to the field of autonomous vehicles are discussed. These applications are structured according to the main Soft Computing technique involved, concretely, fuzzy logic, genetic algorithms and neural networks.

2.1 Fuzzy Logic in Autonomous Driving

Due to the ability of representing expert knowledge in the form of simple and legible rules, fuzzy logic has been broadly used in the field of autonomous vehicles, mainly in the development of control algorithms. In this case, control of a complex nonlinear system can be expressed in the form of a set of simple fuzzy rules (e.g. if speed is too high, then press brake). Fuzzy based methods are specially indicated when we try to emulate human control actions, such as human car driving [59].

Examples of the use of fuzzy rules for the control of the elements of an autonomous vehicle can be found in the recent literature in large amount of examples. In [64], Rodriguez-Castao et al. used a Takagi-Sugeno-Kang fuzzy system for GPS based autonomous navigation of heavy vehicles at high speed. Other example can be found in [22], where Faddel et al. used a fuzzy system to manage the controller for electric vehicle charging. The steering control of autonomous vehicles has received important attention from researchers in fuzzy logic, examples of such interest are [5], where the authors proposed an autonomous platooning system for trucks, including steering control, in order to increase in traffic capacity. Motion planning has been another topic where fuzzy logic has been applied. For instance, in [37], Kala and Warwick implemented decision making over autonomous vehicle maneuvering. Pedals control by fuzzy logic has been also tackled in several works, as [51] or [34].

2.2 Metaheuristics in Autonomous Driving

Metaheuristics, specially genetic and evolutionary algorithms, have been extensively used, as optimization methods in the field of autonomous driving. Their main objective has been the tuning of control systems, decision making and improvement of the efficiency of the overall traffic. In conjunction with fuzzy logic, as the way of representing knowledge, these algorithms have been extensively used for optimizing the distribution of membership functions, rule base or both, under the paradigm of the genetic fuzzy systems.

In [19], Du et al. used a genetic algorithm to optimize a model predictive controller for the simultaneous control of the steering and pedals of an autonomous vehicle, taking the comfort of the passengers as an objective to optimize. A path planning and scheduling method for fleets of autonomous vehicles was proposed in [79]; in this work, Xidias et al. used a genetic algorithm to obtain a near optimum solution for a problem resulting of combining both (planning and scheduling) ones. A multi-objective genetic algorithm was used by Onieva et al. in [52] in order to generate speed profiles for autonomous vehicles to follow in order to cross an intersection where no cooperation among vehicles is possible. Finally, parking trajectories have also been candidates to be optimized by evolutionary algorithms [91].

2.3 Neural Networks in Autonomous Driving

Neural networks (NNs) provide a set of qualities that makes them extremely precise for the representation of complex non-linear systems. They have been used in the field of autonomous vehicles, in one hand, for the control of the actuators of the vehicle, but also for the processing of the high amount of data received by the vehicle, in particular under the field of computer vision.

Recently, with the explosion of deep learning paradigm [66], researchers have defined a new framework where all the information available is used to train models, which are increasing in accuracy as new elements are fed into the NN [15]. They have been applied by a large number of researchers to the processing of visual information captured by autonomous vehicles. In [36], Jia et al. used deep neural networks to provide precise obstacle detection in front of an autonomous vehicle, as well as to segment obstacles and infer their depths. A convolutional NN is used in [88] by Yang et al. to classify roads signs in a hierarchical way, obtaining both the subclasses within each superclass exposed in a picture. In combination with fuzzy logic, the work by Barman et al. [7] presented a fuzzy-NN to guide an unmanned vehicle for maintaining traffic rules to reach its goal and avoid obstacles. Examples of control of actuators at a low level by means of NNs can be found in [17] for lateral, and [58] for the longitudinal control of autonomous vehicles.

3 Soft Computing in Traffic State Prediction

According to the Eurobarometer 2014 about the quality of the transportation, the preferred mode of transport in a typical day is the car, well above from urban public transport. This issue, added to the fact that vehicles per capita have been increased in the last 10 years, has raised the efficiency of the transport to the level of fundamental condition, especially in big cities. For these reasons, road trips are a key point inside ITS, due to the importance in daily life not only for people but also for transportation companies. Inside this subject, one field where different techniques are being used during the last years with a high impact and reliable performance is the prediction of the traffic state in freeway and urban scenarios. One of the principal challenges in this field is to predict, with a certain level of confidence, possible traffic jams in a short-term horizon. The principal advantage of the successful prediction of traffic jams is the adaptation of decision making in the exact moment when different events that may affect traffic, as for example accidents, occur. Another advantage is the capability of calculating not only the travel time but also of planning the route to follow before its beginning. If the user knows the probability of finding a traffic jam in its route, he/she can avoid it by changing the route before or even during the journey. In a general way, the successful prediction of traffic jams can lead to the decrease of travel time, the reduction of CO_2 emissions as well as fuel consumption, or the decrease of acoustic contamination in urban and freeway environments.

Following the same guidelines of the previous section, some relevant applications of Soft Computing in traffic state prediction are reviewed. Concretely, the application of the three components of Soft Computing most commonly used in this topic as neural networks, fuzzy logic and probabilistic reasoning.

3.1 Neural Networks in Traffic State Prediction

In the last years, traffic congestion prediction is one of the fields where NNs have been widely used, as can be seen in literature. For example, in [42], Kumar et al. applied a NN to predict traffic congestion using historical traffic data. Volume, speed, density, and both time and day of the week were used as input variables. The model was validated using rural highway traffic. Another case was presented in [50] by Oh et al., where Gaussian mixture model clustering is combined with a NN to create an urban traffic flow prediction system. The system forecasts traffic flow by combining road geographical and environmental factors with traffic flow properties obtained by the use of detectors. Another type of NNs, called Back Propagation NN (BPNN) is used to forecast campus traffic congestion level in [90]. The results are compared with a Markov model, and the BPNN achieved higher accuracy and more stable performance.

In [41], Koesdwiady et al. used a deep belief networks to enhance prediction accuracy using weather conditions. The study had two objectives: to investigate a correlation between weather parameters and traffic flow, and to improve traffic flow prediction accuracy. The data used for this paper was originated from San Francisco Bay area of California. A Big Data-based framework was adopted in [68] by Soua et al. to address the problem of short-term traffic flow prediction. Deep belief networks are used to independently predict traffic flow using historical traffic flow and weather data, and event-based data.

3.2 Probabilistic Reasoning in Traffic State Prediction

Authors of [4] presented an hybrid approach of parametric and nonparametric methods such as an ensemble of Kalman Filter and NNs in order to improve the travel time prediction of journeys that starts from 15 to 30 min in the future. Kalman Filter was combined with ARIMA in [80], where the ARIMA model is built using historical traffic data. After that, the model is integrated with a Kalman Filter to construct a road traffic state prediction algorithm. Four road segments in Beijing were adopted for the case studies accomplished.

In [24], Fusco et al. hybridized Bayesian networks and NN to create short-term prediction models using as data the link speeds recorded on the metropolitan area of Rome during 7 months. Other example where Bayesian networks are applied to short-term traffic prediction was presented in [81]. In the mentioned paper, traffic flow is predicted using a Bayesian multivariate adaptive-regression splines model. Data is collected from a series of observation stations along the freeway Interstate 205 in Portland, USA, and used to evaluate the performance of the model. Results were compared with different methods, as ARIMA, seasonal ARIMA, and a Kernel method Support Vector Regression.

3.3 Fuzzy Logic in Traffic State Prediction

As mentioned above, fuzzy logic allows to process imprecise information using IF-THEN rules, which helps to the interpretation of the final model. One of the most used and known types of fuzzy systems are Fuzzy Rule Based Systems (FRBS), which can be divided into Mamdani and Takagi-Sugeno-Kang (TSK) systems. Besides, another kind of systems, based in the previous ones, are called Hierarchical FRBS (HFRBS). This class of systems counts with several FRBSs, which are joined in a way that the output of one of them is connected to the input of another one. Depending of the structure of the hierarchy, those systems can be divided into parallel, serial, and hybrid [9].

In traffic congestion prediction, those systems have been used in [93, 94] to develop a congestion prediction system employing a large number of input variables. In these papers, a Steady-State GA is applied to tune the different parts of the FRBSs. A related work is presented in [45], where Lopez-Garcia et al. used a hybrid algorithm that combines GA and Cross Entropy method to tune a HFRBS in order to predict congestion in a freeway in California with time horizons of 5, 15, and 30 min. An extension of that work is presented in [46], where state-of-the-art techniques are compared with the results obtained by the tuned HFRBSs in different traffic congestion datasets. Finally, another interested paper in this topic was presented in [53] Onieva et al. where the authors compare the performance of several Evolutionary Fuzzy and Crisp Rule Based methods for traffic congestion prediction.

4 Soft Computing in Vehicle Route Planning

Other field in which soft computing techniques have demonstrated an outstanding performance is vehicle route planning or vehicle routing problems. Nowadays, route planning is a widely studied field in which the most used and well-known problems are the Traveling Salesman Problem (TSP) [43], and the Vehicle Routing Problem (VRP) [44], being the focus of a big amount of studies in the literature.

The reasons for the importance and popularity of this kind of problems are both scientific and social. On the one hand, most of the problems arising in this field have a great complexity since they belong to NP-Hard class. For this reason, their resolution supposes a major challenge for the scientific community. On the other hand, routing problems are usually built to address real world situations related to logistics, transportation, electronics, robotics, etc.

The first part of this section is focused on metaheuristics, whose application in the resolution of these optimization problems has been very successful. The second part revolves around the use of fuzzy logic in routing problems, describing some relevant works published in the last years.

4.1 Metaheuristics in Vehicle Route Planning

Metaheuristics have been widely used for the solving of routing problems in the last decades, becoming the state-of-the-art in the resolution of many of the variants of these problems. One of the first metaheuristics applied in this context was Simulated Annealing (SA) [74]. For example, in [47], Malek et al. presented a serial and parallel SA for solving the TSP. Other example of the application of this technique for route planning is the work published by Chiang and Rusell [16], in which the VRP with Time Windows is solved using a SA. More recently, Baos et al. developed a parallel variant of SA, called Multiple Temperature Pareto SA in [6], to also solve the VRP with Time Windows with very successful results. Another well-known stochastic local search, Tabu Search (TS), has been also frequently used for solving route planning problems. A recent work on this topic is the one presented by Escobar et al. in 2014 [21], in which they proposed a hybrid granular TS for tackling the challenging Multi-Depot VRP. Briefly explained, the proposed method considers different neighborhoods and diversification strategies, with the aim of improving the initial solution obtained by a hybrid procedure. The Variable Neighborhood Search (VNS) has also demonstrated its efficiency in this area. An interesting example is the work presented in [14], in which Carrabs et al. proposed a VNS for solving a multi-attribute version of the TSP: a Pickup and Delivery TSP with LIFO Loading. More concretely, the authors of this paper introduce three new local search operators, which are then embedded within a VNS. In a more recent publication, Sarasola et al. [65] developed a VNS for facing a stochastic and dynamic VRP. This version of the VRP contemplates two different features. The first one is stochastic demand, which is only revealed when the vehicle arrives at the customer location. The second feature is the dynamic request, meaning that new orders from previously unknown customers can be received and scheduled over time.

Furthermore, evolutionary methods have also shown a great performance for this sort of problems, being Genetic Algorithms (GA) one of the most successful ones. The work presented by Vidal et al. in 2013 is an example of this fact [77]. In this research, a hybrid genetic algorithm with adaptive diversity management is implemented for tackling the VRP with time windows. Another example is the survey paper published by Karakatič and Podgorelec in 2015 [39], which collects some of the most important works focused on the application of the GA to the multi-depot VRP.

Additionally, since the appearance of GA in the early 1970s, a wide variety of nature-inspired metaheuristics have also appeared in literature. Some of these recently proposed methods are the Firefly Algorithm (FA) and the Bat Algorithm (BA). The FA was proposed by Yang in 2008 [84]. This meta-heuristic has been applied to a wide range of optimization fields and problems since its proposal [87], and it has also shown a promising performance for routing problems. In [35], for example, Jati and Suyanto presented the first application of the FA for solving the TSP. In order to do that, authors adapt the FA, which was firstly proposed for tackling continuous problems, providing it with an evolutionary and discrete behavior. Another interesting example of application is the one presented in [3] by Alinaghian and Naderipour, in which a hybrid version of the FA is proposed to solve a time-dependent VRP with multi-alternative graph, in order to reduce the fuel consumption. The developed hybrid version of the FA is a Gaussian Firefly Algorithm. The most interesting part of this paper is the real-world use case that authors presented, focused on a distribution company, established in Esfahan, Iran. Additionally, in [56] Osaba et al. also shown that the FA is able to face complex routing problem, such as the asymmetric and clustered VRP with simultaneous pickup and deliveries, variable costs and forbidden paths. Finally, in [54], the same authors presented a evolutionary discrete FA with a novel operator to deal with VRP with time windows with successful results.

Regarding the other nature-inspired method mentioned above, the BA, it was proposed by Yang in 2010 [85]. As can be read in several surveys [86], the BA has been successfully applied to different optimization fields and problems since its proposal. Focusing in routing problems, several recent papers have shown that the BA is a promising technique in vehicle route planning. For example, in [70], Taha et al. presented an adapted version of this algorithm for solving the well-known Capacitated VRP. The Adapted BA developed in that study allows a large diversity of the population and a balance between global and local search. Zhou et al. addressed the same problem in [95]. In that paper a hybrid BA with path relinking is described. This approach is constructed based on the framework of the continuous BA, in which the greedy randomized adaptive search procedure and path relinking are effectively integrated. Additionally, with the aim of improving the performance of the technique, the random subsequences and single-point local search are operated with certain probability. In [55], Osaba et al. presented an improved adaptation of the

BA for addressing both symmetric and asymmetric TSP. The results shown that the improved version of BA could obtain promising results, in comparison with some reference techniques, such as an evolutionary simulated annealing, a genetic algorithm, a distributed genetic algorithm or an imperialist competitive algorithm.

We want to highlight that the meta-heuristics referenced in this section form a small part of all different approaches that can be found in current literature. We are aware that many other interesting and efficient techniques are available in the scientific community, such as the Harmony Search [27], or Gravitational Search [61, 62], which also show a good performance when they are applied to routing problems. Additionally, many additional classic methods have also shown a great performance for this kind of problems, such as the Particle Swarm Optimization [89], the Ant Colony Optimization [63] or Large Neighborhood Search [60].

4.2 Fuzzy Logic in Vehicle Route Planning

The use of fuzzy systems is also an important topic in the field of vehicle routing problems. In real situations, these problems are susceptible to suffer imprecision or uncertainty in their data. Many works in literature show that one of the most successful ways to tackle with this uncertainty and imprecision in the information available when solving vehicle route planning problems, it is the use of fuzzy logic.

In this way, we can find interesting studies in the literature, such as the one presented in [29], in which Ghannadpour et al. proposed a multi-objective dynamic vehicle routing problem with fuzzy time windows. In this research, authors not only describe the problem, but also the main real-world applications that it could have. The constraints related with travel times and user satisfaction are some of the most subject to uncertainty and imprecision. Apart from the previous work, this type of imprecision is modelled in other studies such as the one presented by Tang et al. in [71], in which a VRP with fuzzy time windows is proposed and solved using a two-stage algorithm which decomposes the problem into two subproblems. An additional example of this trend is the work proposed in [28], in which a multi-objective dynamic VRP with fuzzy travel times and customers' satisfaction level is presented. Specifically, the customers' satisfaction level is considered in the route planning of vehicles by using the concept of fuzzy time windows. Additionally, the dynamic solving strategy proposed is based on a genetic algorithm, and its performance is evaluated on various test problems generalized from a set of static instances in the literature. Other interesting application of fuzzy logic for vehicle routing problems are the works presented in [12], where Brito et al. proposed a variant of the close-open VRP with fuzzy time windows and fuzzy vehicle capacity, and [72], where Torres et al. solved a variant of the Truck and Trailer Routing Problem where the imprecision in the capacity of the truck and the trailers is modeled by fuzzy logic.

5 Soft Computing in Vehicular Ad-Hoc Networks

Vehicular Ad-Hoc Networks (VANETs) are communication networks in which the nodes are vehicles [32]. This field has attracted the attention of the scientific community, automobile industry and institutions worldwide because of the huge number of innovative applications they can enable [57]. Among the areas of application, some of the most relevant are: security (warnings about emergency break, collision at intersection, line shift, etc.); leisure and entertainment (multimedia content download, nearby points of interest, etc.); traffic management (virtual traffic lights, limited access zones, electronic tolls, etc.); and driver assistance (remote diagnosis; efficient and eco-driving; etc.). The communications that take place within VANETs can be classified in Infrastructure-to-Infrastructure (I2I), Infrastructure-to-Vehicle (I2V), Vehicle-to-Infrastructure (V2I) and Vehicle-to-Vehicle (V2V) depending on which agent is the transmitter and the receptor.

VANETs are an especially complex environment because of the high dynamism in the movement of the vehicles, communication failures, different driver profiles, high variability of nodes and interconnections, etc. For this reason, data handled in VANETs is subject to imprecision, uncertainty and vagueness which make it an excellent field for the application of Soft Computing techniques. This section is devoted to review some of these applications. Concretely, the next subsections will show, in this order, applications in VANETs of three components of Soft Computing as metaheuristics, fuzzy sets and neural networks.

5.1 Metaheuristics in Vehicular Ad-Hoc Networks

In the context of VANETs, several complex optimization problems appear, in which metaheuristics have shown to be an excellent tool to solve them. One of the first works in which metaheuristics were applied to VANETs can be found in [26]. Here, Garcia-Nieto and co-authors proposed the used of these techniques to optimize the File Transfer Protocol Configuration. Concretely, they optimized the parameters of the Vehicular Data Transfer Protocol, which operate over the transport layer protocol of VANETs, in order to allow end-to-end communications. To address this problem, they test five different metaheuristics over two scenarios that simulated urban and highway environments. The authors concluded that the metaheuristics reduced the transmission time in a 19 and 25.43% in urban and highway scenarios, respectively, when they were compared to the configuration provided by a human expert. A similar approach was followed by the same authors in [73]. In this case, the five metaheuristics mentioned before were employed to find the optimal configuration of the Open Link State Routing protocol for VANETs. The results showed again a significant improvement in terms of Packet Delivery Ratio (PDR), network routing load and End-to-End Delay (E2ED) in comparison to standard and expert configurations.

Another optimization problem from VANETs that has been addressed with metaheuristics is multi-cast routing. Souza et al. presented in [69] a tree based multicast routing protocol called MAV-AODV. Here, the Ant Colony Optimization's pheromone mechanism is used to establish a quality measure of the stability of the routes. The new method was tested over a simulated Manhattan scenario and compared with the MAODV protocol. MAV-AODV obtained a better performance than MAODV in terms of E2ED, overhead and PDR. Other example of routing protocol inspired in metaheuristics was presented in [10]. This unicast and multipath protocol, called HyBR, used two types of routing procedures: a topology-based and a geographic-based routing procedure for high and low density scenarios, respectively. The first one was inspired in the working of bee swarm optimization whereas the second one used a genetic algorithm to optimize the route between the origin and the destiny. The experimentation was done over high density and a low density scenarios and the performance measures considered were the average E2ED, PDR and normalized overhead load. HyBR outperformed AODV and geography-based routing protocol (GPSR) in the first two measures but not in the last one.

A more recent application of metaheuristics in VANETs is given by Masegosa et al. in [48]. This work is focused on information dissemination from a central server to vehicles by means of Virtual Infrastructures (VIs). The selection of the nodes of the VI is modelled as a covering location problem and it is solved by means of a genetic algorithm. The main challenge for metaheuristics in this environment was the short response time imposed by the latency requirements of some VANET's applications. The experimentation over a real scenario with 45 vehicles indicated that the proposal outperforms another state-of-art method based on a deterministic greedy strategy, called NAVI.

5.2 *Fuzzy Logic in Vehicular Ad-Hoc Networks*

Fuzzy set theory has been also applied in different areas of VANETs. For example, in [30], Abdel Hafeez et al. presented a Cluster Head (CH) selection mechanism that made use of a fuzzy inference system. CH selection is one of the main challenges of cluster-based medium access control protocols, whose aim is to improve the access and capacity of the network among other aspects. The previous mechanism elected the CHs dynamically and taking into account a stability criteria. The fuzzy inference system was used to predict the future position and speed of all cluster members using as input the inter-vehicle distance and speed. The procedure proposed outperforms CMCP [31] and APROVE [67] protocols.

Another fuzzy inference system was presented in [78] to design a multi-hop broadcast protocol, named FUZZYBR. In a more specific way, the fuzzy inference system was employed to select the relay nodes considering variables with a high degree of imprecision and uncertainty as the inter-vehicle distance, mobility and signal strength. The evaluation of the methods was done over simulated freeway and street scenarios, and the proposal was compared with other broadcast protocols as

Flooding, Weighted persistence, MPR and Enhance MPR. The results confirmed a significant performance advantage of FUZZYBR over the mentioned protocols in terms of PDR, E2ED and number of messages per data package. In [25], Galaviz-Mosqueda et al. utilized FUZZYBR and another multi-hop broadcast protocol based on fuzzy inference systems, and called RLMB, to test the use of genetic fuzzy systems. The motivation of the authors was to adjust the membership functions of the fuzzy rules of the two protocols by means of a genetic algorithm, in order to obtain a better performance than the one with expert tuning. The results confirmed their hypothesis, and the two versions of FUZZYBR and RLMB automatically tuned with the genetic algorithm significantly improved the performance of the counterparts heuristically configured by humans.

Fuzzy control was also applied in VANETs to adapt beaconing rate to the changes in traffic density that usually occur along time. This mechanism, called ABR, was proposed in [96] where the authors developed a method in which a rule-based system adapted the frequency of beacon broadcasting taking into account the percentage of vehicles traveling in the same direction and the emergency status of vehicles. The simulations showed how this method reduced the beaconing load at the expense of cooperative awareness between vehicles.

5.3 Neural Networks in VANETs

One of the first works that suggested the application of NNs to VANETs can be found in [40]. In this paper, the authors aimed at demonstrating the benefits of VANETs on traffic safety, and concretely for designing an Accident Prevention Application (APA). To this end, they proposed, in a first stage, the use of a Markov Reward Process to estimate the expected time until an accident will happen, taking into account the traffic states observed so far; and in a second stage, the use of NNs to make these estimations when there are unobserved traffics situation. To this end, the NNs should have been trained with known pairs (state, expected time). The authors claimed that, in this way, they provided the basis for the analysis of VANETs and their impact on traffic safety.

In a more recent paper [83], Yang et al. combined ANN and VANETs to develop a short-term average-speed forecast and adjustment approach to improve gas consumption, decrease CO_2 emissions and reduce travel time. In the proposed method, a Traffic Information Center (TIC) collected average speed from vehicles and road side sensors through VANETs. Then, the TIC trained a NN with average speed, weather information and traffic flow to predict the average speed. The predicted speed was then sent to the CH that adjusted the prediction according to the observed speed. The simulations done showed an important improvement in the accuracy of the average-speed predictions when the system was compared versus a hybrid approach.

Another important issue in VANETs handled with NNs is security and vulnerability, given that VANETs are even more exposed than other similar networks. A good example of this application can be found in [2], where the authors employed

NNs to build an Intrusion Detection System (IDS) to prevent Denial of Services (DoS) attacks. Concretely, the aim of the NN was the real-time detection of malicious vehicles in order to isolate them from the network. With this purpose in mind, the authors generated data for normal and malicious vehicles through simulations. From this data, they extracted relevant features and a pre-processed dataset that it was used to train the NN. The experimentation showed that the system obtained an error rate of 2.05%, confirming its effectiveness. Other example of the application of NNs for security in VANETs has been recently presented in [38]. The authors of this work developed a Deep NN (DNN) for an IDS to secure in-vehicular networks that use the CAN protocol. Concretely, the proposed IDS considered a scenario in which malicious data-package are injected into the in-vehicle CAN bus. The DNN was trained with labeled (i.e. normal or attack) and preprocessed CAN packets to extract features that model the statistical behavior of the network. In the detection phase, each CAN packet was pre-processed and passed to the trained DNN to make the binary decision. The experimentation demonstrated that the approach obtained a 98% detection ratio in real-time response to attacks.

6 Discussions

In this book chapter we have presented an overview of application of SC to four important areas of ITS: autonomous driving, traffic state prediction, vehicle route planning and VANETs. Our overview has shown that SC techniques are an effective and efficient framework to deal with many of the problems that arise in those areas and therefore, to develop better performing ITS.

The main reasons behind the success of SC in the four ITS areas aforementioned is associated with the recent trend in ITS to follow a data-driven approach; and the inherent tolerance of SC techniques to deal with the imprecision, uncertainty and vagueness, omnipresent in the information handled in this complex environments, and their ability to provide cost-effective solutions.

To finish, we would like to point out that the emergence of new ITS technologies such as autonomous cars, electric vehicles, more advance VANETs or Unmanned Aerial Vehicles will probably boost a shift paradigm, along the next decade, in the way in which goods and persons are transported nowadays. SC plays and will play a major role in this shift so we augur a great future for the application and development of SC techniques in this field.

Acknowledgements This work has been supported by the research projects TEC2013-45585-C2-2-R and TIN2014-56042-JIN from the Spanish Ministry of Economy and Competitiveness, and TIMON project which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 636220.

References

1. Road safety in the European union trends, statistics and main challenges. Technical report, European Commission: General Mobility and Transport, Unit C.4 Road Safety (2015)
2. Alheeti, K.M.A., Gruebler, A., McDonald-Maier, K.D.: An intrusion detection system against malicious attacks on the communication network of driverless cars. In: 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), pp. 916–921 (2015)
3. Alinaghian, M., Naderipour, M.: A novel comprehensive macroscopic model for time-dependent vehicle routing problem with multi-alternative graph to reduce fuel consumption: a case study. Comput. Ind. Eng. **99**, 210–222 (2016)
4. Allström, A., Ekström, J., Gundlegård, D., Ringdahl, R., Rydergren, C., Bayen, A., Patire, A.: Hybrid approach for short-term traffic state and travel time prediction on highways. Trans. Res. Rec. **2554**, 60–68 (2016)
5. Ario, T., Sugimachi, T., Fukao, T., Kawashima, H.: Evaluation of fuzzy inference-based self-tuning of steering control gains for heavy-duty trucks. Int. J. Intell. Trans. Syst. Res. **14**(2), 92–100 (2016)
6. Baños, R., Ortega, J., Gil, C., Fernández, A., de Toro, F.: A simulated annealing-based parallel multi-objective approach to vehicle routing problems with time windows. Expert Syst. Appl. **40**(5), 1696–1707 (2013)
7. Barman, B., Kanjilal, R., Mukhopadhyay, A.: Neuro-fuzzy controller design to navigate unmanned vehicle with construction of traffic rules to avoid obstacles. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **24**(3), 433–449 (2016)
8. Belk, R.: You are what you can access: sharing and collaborative consumption online. J. Bus. Res. **67**(8), 1595–1600 (2014)
9. Bentez, A., Casillas, J.: Multi-objective genetic learning of serial hierarchical fuzzy systems for large-scale problems. Soft Comput. **17**(1), 165–194 (2013)
10. Bitam, S., Mellouk, A., Zeadally, S.: HyBR: A hybrid bio-inspired bee swarm routing protocol for safety applications in vehicular ad hoc networks (VANETs). J. Syst. Archit. **59**(10), 953–967 (2013)
11. Bonnefon, J.F., Shariff, A., Rahwan, I.: Autonomous vehicles need experimental ethics: are we ready for utilitarian cars? arXiv preprint. arXiv:1510.03346 (2015)
12. Brito, J., Martínez, F.J., Moreno, J.A., Verdegay, J.L.: An ACO hybrid metaheuristic for close-open vehicle routing problems with time windows and fuzzy constraints. Appl. Soft Comput. **32**, 154–163 (2015)
13. Brynjolfsson, E., McAfee, A.: Race Against the Machine. Digital Frontier, Lexington, MA (2011)
14. Carrabs, F., Cordeau, J.F., Laporte, G.: Variable neighborhood search for the pickup and delivery traveling salesman problem with lifo loading. Inf. J. Comput. **19**(4), 618–632 (2007)
15. Chen, C., Seff, A., Kornhauser, A., Xiao, J.: Deepdriving: learning affordance for direct perception in autonomous driving. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2722–2730 (2015)
16. Chiang, W.C., Russell, R.A.: Simulated annealing metaheuristics for the vehicle routing problem with time windows. Ann. Oper. Res. **63**(1), 3–27 (1996)
17. Chouraqui, S., Selma, B.: Unmanned vehicle trajectory tracking by neural networks. Int. Arab J. Inf. Technol. **13**(3), 272–275 (2016)
18. Dimitrakopoulos, G., Demestichas, P.: Intelligent transportation systems. IEEE Veh. Technol. Mag. **5**(1), 77–84 (2010)
19. Du, X., Htet, K., Tan, K.: Development of a genetic-algorithm-based nonlinear model predictive control scheme on velocity and steering of autonomous vehicles. IEEE Trans. Ind. Electron. **63**(11), 6970–6977 (2016)
20. ERTRAC: Multi-annual implementation plan for Horizon 2020. Tech. rep. (2013). http://www.ertrac.org/uploads/documentsearch/id20/ertrac-map-h2020_67.pdf
21. Escobar, J.W., Linfati, R., Toth, P., Baldoquin, M.G.: A hybrid granular tabu search algorithm for the multi-depot vehicle routing problem. J. Heuristics **20**(5), 483–509 (2014)

22. Faddel, S., Mohamed, A., Mohammed, O.: Fuzzy logic-based autonomous controller for electric vehicles charging under different conditions in residential distribution systems. Electr. Power Syst. Res. **148**, 48–58 (2017)
23. Figueiredo, L., Jesus, I., Machado, J., Ferreira, J., Martins de Carvalho, J.: Towards the development of intelligent transportation systems. In: Proceedings of the 2001 IEEE Intelligent Transportation Systems, pp. 1206–1211 (2001)
24. Fusco, G., Colombaroni, C., Isaenko, N.: Comparative analysis of implicit models for real-time short-term traffic predictions. IET Intell. Trans. Syst. **10**(4), 270–278 (2016)
25. Galaviz-Mosqueda, A., Villarreal-Reyes, S., Galeana-Zapien, H., Rubio-Loyola, J., Rivera-Rodriguez, R.: Genetic tuning of fuzzy rule-based systems for multi-hop broadcast protocols for VANETs. Telecommun. Syst. **63**(3), 399–420 (2016)
26. García-Nieto, J., Toutouh, J., Alba, E.: Automatic tuning of communication protocols for vehicular ad hoc networks using metaheuristics. Eng. Appl. Artif. Intell. **23**(5), 795–805 (2010)
27. Geem, Z.W., Kim, J.H., Loganathan, G.: A new heuristic optimization algorithm: harmony search. Simulation **76**(2), 60–68 (2001)
28. Ghannadpour, S.F., Noori, S., Tavakkoli-Moghaddam, R.: Multiobjective dynamic vehicle routing problem with fuzzy travel times and customers satisfaction in supply chain management. IEEE Trans. Eng. Manag. **60**(4), 777–790 (2013)
29. Ghannadpour, S.F., Noori, S., Tavakkoli-Moghaddam, R., Ghoseiri, K.: A multi-objective dynamic vehicle routing problem with fuzzy time windows: model, solution and application. Appl. Soft Comput. **14**, 504–527 (2014)
30. Hafeez, K.A., Zhao, L., Liao, Z., Ma, B.N.W.: A fuzzy-logic-based cluster head selection algorithm in VANETs. In: 2012 IEEE International Conference on Communications (ICC), pp. 203–207 (2012)
31. Hang, Su, Zhang, Xi: Clustering-based multichannel MAC protocols for QoS provisionings over vehicular ad hoc networks. IEEE Trans. Veh. Technol. **56**(6), 3309–3323 (2007)
32. Hartenstein, H., Laberteaux, K.P.: A tutorial survey on vehicular ad hoc networks. IEEE Commun. Mag. **46**(6), 164–171 (2008)
33. Hatcher, G., Burnier, C., Greer, E., Hardesty, D., Hicks, D., Jacobi, A., Lowrance, C., Mercer, M.: Intelligent transportation systems benefits, costs, and lessons learned: 2014 update report. In: Technical report, U.S. Department of Transportation, ITS Joint Program Office (2014). https://trid.trb.org/view.aspx?id=1334694
34. Huang, W., Wu, Q., Ma, Y.L., Chu, X.M.: Design of speed controller of small intelligent vehicle based on visual navigation. J. Wuhan Univ. Technol. **32**(6), 103–106 (2010)
35. Jati, G.K., Suyanto: Evolutionary discrete firefly algorithm for travelling salesman problem. In: Proceedings of the Adaptive and Intelligent Systems: Second International Conference, pp. 393–403 (2011)
36. Jia, B., Feng, W., Zhu, M.: Obstacle detection in single images with deep neural networks. Signal, Image Video Process. **10**(6), 1033–1040 (2016)
37. Kala, R., Warwick, K.: Reactive planning of autonomous vehicles for traffic scenarios. Electronics **4**(4), 739–762 (2015)
38. Kang, M.J., Kang, J.W., Wang, X., Larochelle, H., Vincent, P., Bengio, S.: Intrusion detection system using deep neural network for in-vehicle network security. PLOS ONE **11**(6), e0155, 781 (2016)
39. Karakatič, S., Podgorelec, V.: A survey of genetic algorithms for solving multi depot vehicle routing problem. Appl. Soft Comput. **27**, 519–532 (2015)
40. Killat, M., Hartenstein, H.: Vehicular ad hoc networks: how to show the impact on traffic safety? In: 2007 IEEE 65th Vehicular Technology Conference—VTC2007, Spring, pp. 659–663 (2007)
41. Koesdwiady, A., Soua, R., Karray, F.: Improving traffic flow prediction with weather information in connected cars: a deep learning approach. IEEE Trans. Veh. Technol. **65**(12), 9508–9517 (2016)
42. Kumar, K., Parida, M., Katiyar, V.: Short term traffic flow prediction in heterogeneous condition using artificial neural network. Transport **30**(4), 397–405 (2015)

43. Laporte, G.: The traveling salesman problem: an overview of exact and approximate algorithms. Eur. J. Oper. Res. **59**(2), 231–247 (1992)
44. Laporte, G.: The vehicle routing problem: an overview of exact and approximate algorithms. Eur. J. Oper. Res. **59**(3), 345–358 (1992)
45. Lopez-Garcia, P., Onieva, E., Osaba, E., Masegosa, A.D., Perallos, A.: A hybrid method for short-term traffic congestion forecasting using genetic algorithms and cross entropy. IEEE Trans. Intell. Trans. Syst. **17**(2), 557–569 (2016)
46. Lopez-Garcia, P., Osaba, E., Onieva, E., Masegosa, A.D., Perallos, A.: Short-term traffic congestion forecasting using hybrid metaheuristics and rule-based methods: a comparative study. In: Luaces, O., Gámez, J.A., Barrenechea, E., Troncoso, A., Galar, M., Quintián, H., Corchado, E. (eds.) Advances in Artificial Intelligence: 17th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2016, pp. 290–299. Springer International Publishing (2016)
47. Malek, M., Guruswamy, M., Pandya, M., Owens, H.: Serial and parallel simulated annealing and tabu search algorithms for the traveling salesman problem. Ann. Oper. Res. **21**(1), 59–84 (1989)
48. Masegosa, A.D., de la Iglesia, I., Hernandez-Jayo, U., Diez, L.E., Bahillo, A., Onieva, E.: Solutions based on soft computing for the sustainability and climate change. In: A New Approach for Information Dissemination in VANETs Based on Covering Location and Metaheuristics, pp. 1–24. Springer International Publishing (2017). In press
49. OECD: Improving Reliability on Surface Transport Networks. OECD Publishing (2010). URL http://www.oecd-ilibrary.org/transport/improving-reliability-on-surface-transport-networks_9789282102428-en
50. Oh, S.D., Kim, Y.J., Hong, J.S.: Urban traffic flow prediction system using a multifactor pattern recognition model. IEEE Trans. Intell. Trans. Syst. **16**(5), 2744–2755 (2015)
51. Onieva, E., Godoy, J., Villagrá, J., Milanés, V., Pérez, J.: On-line learning of a fuzzy controller for a precise vehicle cruise control system. Expert Syst. Appl. **40**(4), 1046–1053 (2013)
52. Onieva, E., Hernandez-Jayo, U., Osaba, E., Perallos, A., Zhang, X.: A multi-objective evolutionary algorithm for the tuning of fuzzy rule bases for uncoordinated intersections in autonomous driving. Inf. Sci. **321**, 14–30 (2015)
53. Onieva, E., Lopez-Garcia, P., Masegosa, A.D., Osaba, E., Perallos, A.: A comparative study on the performance of evolutionary fuzzy and crisp rule based classification methods in congestion prediction. Trans. Res. Proced. **14**, 4458–4467 (2016)
54. Osaba, E., Carballedo, R., Yang, X.S., Diaz, F.: An Evolutionary Discrete Firefly Algorithm with Novel Operators for Solving the Vehicle Routing Problem with Time Windows, pp. 21–41. Springer International Publishing, Cham (2016)
55. Osaba, E., Yang, X.S., Diaz, F., Lopez-Garcia, P., Carballedo, R.: An improved discrete bat algorithm for symmetric and asymmetric traveling salesman problems. Eng. Appl. Artif. Intell. **48**, 59–71 (2016)
56. Osaba, E., Yang, X.S., Diaz, F., Onieva, E., Masegosa, A.D., Perallos, A.: A discrete firefly algorithm to solve a rich vehicle routing problem modelling a newspaper distribution system with recycling policy. Soft Comput. 1–14 (2016)
57. Papadimitratos, P., La Fortelle, A., Evenssen, K., Brignolo, R., Cosenza, S.: Vehicular communication systems: enabling technologies, applications, and future outlook on intelligent transportation. IEEE Commun. Mag. **47**(11), 84–95 (2009)
58. Pérez, J., Gajate, A., Milanés, V., Onieva, E., Santos, M.: Design and implementation of a neuro-fuzzy system for longitudinal control of autonomous vehicles. In: International Conference on Fuzzy Systems, pp. 1–6 (2010)
59. Pérez, J., Milanés, V., Godoy, J., Villagrá, J., Onieva, E.: Cooperative controllers for highways based on human experience. Expert Syst. Appl. **40**(4), 1024–1033 (2013)
60. Pisinger, D., Ropke, S.: Large neighborhood search In: Handbook of Metaheuristics, pp. 399–419 (2010)
61. Precup, R.E., David, R.C., Petriu, E.M., Preitl, S., Rădac, M.B.: Fuzzy logic-based adaptive gravitational search algorithm for optimal tuning of fuzzy-controlled servo systems. IET Control Theory Appl. **7**(1), 99–107 (2013)

62. Precup, R.E., David, R.C., Petriu, E.M., Radac, M.B., Preitl, S.: Adaptive GSA-based optimal tuning of PI controlled servo systems with reduced process parametric sensitivity, robust stability and controller robustness. IEEE Trans. Cybern. **44**(11), 1997–2009 (2014)
63. Reed, M., Yiannakou, A., Evering, R.: An ant colony algorithm for the multi-compartment vehicle routing problem. Appl. Soft Comput. **15**, 169–176 (2014)
64. Rodriguez-Castao, A., Heredia, G., Ollero, A.: High-speed autonomous navigation system for heavy vehicles. Appl. Soft Comput. J. **43**, 572–582 (2016)
65. Sarasola, B., Doerner, K.F., Schmid, V., Alba, E.: Variable neighborhood search for the stochastic and dynamic vehicle routing problem. Ann. Oper. Res. **236**(2), 425–461 (2016)
66. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
67. Shea, C., Hassanabadi, B., Valaee, S.: Mobility-based clustering in VANETs using affinity propagation. In: GLOBECOM 2009—2009 IEEE Global Telecommunications Conference, pp. 1–6 (2009)
68. Soua, R., Koesdwiady, A., Karray, F.: Big-data-generated traffic flow prediction using deep learning and dempster-shafer theory. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 3195–3202 (2016)
69. Souza, A.B., Celestino, J., Xavier, F.A., Oliveira, F.D., Patel, A., Latifi, M.: Stable multicast trees based on ant colony optimization for vehicular ad hoc networks. In: The International Conference on Information Networking 2013 (ICOIN), pp. 101–106 (2013)
70. Taha, A., Hachimi, M., Moudden, A.: Adapted bat algorithm for capacitated vehicle routing problem. Int. Rev. Comput. Softw. **10**(6), 610–619 (2015)
71. Tang, J., Pan, Z., Fung, R.Y., Lau, H.: Vehicle routing problem with fuzzy time windows. Fuzzy Sets and Syst. **160**(5), 683–695 (2009)
72. Torres, I., Cruz, C., Verdegay, J.L.: Solving the truck and trailer routing problem with fuzzy constraints. Int. J. Comput. Intell. Syst. **8**(4), 713–724 (2015)
73. Toutouh, J., Garcia-Nieto, J., Alba, E.: Intelligent OLSR routing protocol optimization for VANETs. IEEE Trans. Veh. Technol. **61**(4), 1884–1894 (2012)
74. Van Laarhoven, P.J.M., Aarts, E.H.L.: Simulated annealing. In: Simulated Annealing: Theory and Applications, pp. 7–15 (1987)
75. Verdegay, J.L., Yager, R.R., Bonissone, P.P.: On heuristics as a fundamental constituent of soft computing. Fuzzy Sets and Syst. **159**(7), 846–855 (2008)
76. Verma, M., Manoj, M., Verma, A.: Analysis of aspiration for owning a car among youths in a city of a developing country, India. Trans. Dev. Econ. **3**(1), 7 (2017)
77. Vidal, T., Crainic, T.G., Gendreau, M., Prins, C.: A hybrid genetic algorithm with adaptive diversity management for a large class of vehicle routing problems with time-windows. Comput. Oper. Res. **40**(1), 475–489 (2013)
78. Wu, C., Ohzahata, S., Kato, T.: VANET broadcast protocol based on fuzzy logic and lightweight retransmission mechanism. IEICE Trans. Commun. **E95-B**(2), 415–425 (2012)
79. Xidias, E., Zacharia, P., Nearchou, A.: Path planning and scheduling for a fleet of autonomous vehicles. Robotica **34**(10), 2257–2273 (2016)
80. Xu, D.W., Wang, Y.D., Jia, L.M., Qin, Y., Dong, H.H.: Real-time road traffic state prediction based on arima and kalman filter. Front. Inf. Technol. Electron. Eng. **18**(2), 287–302 (2017)
81. Xu, Y., Kong, Q.J., Klette, R., Liu, Y.: Accurate and interpretable bayesian mars for traffic flow prediction. IEEE Trans. Intell. Trans. Syst. **15**(6), 2457–2469 (2014)
82. Yan, S., Chen, C.Y., Chang, S.C.: A car pooling model and solution method with stochastic vehicle travel times. IEEE Trans. Intell. Trans. Syst. **15**(1), 47–61 (2014)
83. Yang, J.Y., Chou, L.D., Tung, C.F., Huang, S.M., Wang, T.W.: Average-speed forecast and adjustment via VANETs. IEEE Trans. Veh. Technol. **62**(9), 4318–4327 (2013)
84. Yang, X.S.: Nature-inspired metaheuristic algorithms. Luniver press (2008)
85. Yang, X.S.: A new metaheuristic bat-inspired algorithm. In: Nature Inspired Cooperative Strategies for Optimization (NICSO 2010), pp. 65–74 (2010)
86. Yang, X.S., He, X.: Bat algorithm: literature review and applications. Int. J. Bio-Inspir. Comput. **5**(3), 141–149 (2013)

87. Yang, X.S., He, X.: Firefly algorithm: recent advances and applications. Int. J. Swarm Intell. **1**(1), 36–50 (2013)
88. Yang, Y., Luo, H., Xu, H., Wu, F.: Towards real-time traffic sign detection and classification. IEEE Trans. Intell. Trans. Syst. **17**(7), 2022–2031 (2016)
89. Yao, B., Yu, B., Hu, P., Gao, J., Zhang, M.: An improved particle swarm optimization for carton heterogeneous vehicle routing problem with a collection depot. Ann. Oper. Res. **242**(2), 303–320 (2016)
90. Yu, X., Xiong, S., He, Y., Wong, W., Zhao, Y.: Research on campus traffic congestion detection using BP neural network and Markov model. J. Inf. Secur. Appl. **31**, 54–60 (2016)
91. Zhang, D., Li, S., Yang, Q., Liu, L.: Optimization based trajectory planning of parallel parking with multiple constraints. SAE Int. J. Passeng. Cars Electron. Electr. Syst. **8**(2), 413–418 (2015)
92. Zhang, J., Wang, F.Y., Wang, K., Lin, W.H., Xu, X., Chen, C.: Data-driven intelligent transportation systems: a survey. IEEE Trans. Intell. Trans. Syst. **12**(4), 1624–1639 (2011)
93. Zhang, X., Onieva, E., Lee, V., Liu, K.: Congestion prediction by means of fuzzy logic and genetic algorithms. In: Intelligent Transport Systems: Technologies and Applications, pp. 189–205 (2015)
94. Zhang, X., Onieva, E., Perallos, A., Osaba, E., Lee, V.: Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction. Trans. Res. Part C: Emerg. Technol. **43**, 127–142 (2014)
95. Zhou, Y., Luo, Q., Xie, J., Zheng, H.: A hybrid bat algorithm with path relinking for the capacitated vehicle routing problem. In: Yang, X.S., Bekdaş, G., Nigdeli, S.M. (eds.) Metaheuristics and Optimization in Civil Engineering, pp. 255–276. Springer International Publishing, Cham (2016)
96. Zrar Ghafoor, K., AbuBakar, K., van Eenennaam, M., Khokhar, R.H., Gonzalez, A.J.: A fuzzy logic approach to beaconing for vehicular ad hoc networks. Telecommun. Syst. **52**(1), 139–149 (2013)

Fuzzy Cognitive Maps Based Models for Pattern Classification: Advances and Challenges

Gonzalo Nápoles, Maikel Leon Espinosa, Isel Grau, Koen Vanhoof and Rafael Bello

Abstract Fuzzy Cognitive Maps (FCMs) have proven to be a suitable methodology for the design of knowledge-based systems. By combining both uncertainty depiction and cognitive mapping, this technique represents the knowledge of systems that are characterized by ambiguity and complexity. In short, FCMs can be defined as recurrent neural networks that include elements of fuzzy logic during the knowledge engineering phase. While the literature contains many studies claiming how this Soft Computing technique is able to model complex and dynamical systems, we explore another promising research field: *the use of FCMs in solving pattern classification problems*. This is motivated by the transparency of the decision model attached to these cognitive, neural networks. In this chapter, we revise some prominent advances in the area of FCM-based classifiers and open challenges to be confronted.

1 Introduction

In the last years, *Fuzzy Cognitive Maps* (FCMs) [12] have notably increased their popularity within the scientific community. They constitute a suitable tool for the designing of knowledge-based systems, where one of the most relevant characteristics is the interpretability of the network topology. Not many computer science techniques can claim this valuable feature.

From the structural perspective, an FCM can be defined as a fuzzy digraph that describes the underlying behavior of an intelligent system in terms of concepts

G. Nápoles (✉) · K. Vanhoof
Hasselt Universiteit, Agoralaan gebouw D, Diepenbeek, Belgium
e-mail: gonzalo.napoles@uhasselt.be

M. Leon Espinosa
University of Miami, 5250 University Dr, Miami, FL, USA
e-mail: mleon@bus.miami.edu

I. Grau · R. Bello
Central University of Las Villas, Carretera Camajuaní km 5.5, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_5

(i.e., objects, states, variables or entities). Such concepts comprise a precise meaning for the problem domain under analysis and they are connected by signed and weighted edges that denote causal relationships.

The sign and intensity of causal relations involve the quantification of a fuzzy linguistic variable that can be assigned by experts during a knowledge acquisition phase [13]. These elements recurrently interact when updating the activation value of each concept (or simply neuron). In point of fact, an FCM exploits an activation (state) vector by using a rule similar to the standard McCulloch-Pitts scheme [15]. Therefore, the activation value of each neuron is given by the value of the transformed weighted sum that this processing unit receives from connected neurons on the causal network. This activation value actually comprises an interpretable feature for the physical system under investigation. More explicitly, the higher the activation value of a neuron, the stronger its influence (positive or negative) over the connected neural entities. Of course, this influence also depends on the intensity of the causal relations connecting the actual neuron with the other neural processing entities.

FCM-based models can be understood as a kind of *recurrent neural networks* that support backward connections that sometimes form cycles in the causal graph. These backward relations (called feedback) enable the network to handle memory to compute the outputs of the current state and maintain a sort of recurrence to the past processing [6]. During the inference phase, the updating rule is repeated until the system converges to a fixed-point attractor or a maximal number of iterations is reached. The former implies that a hidden pattern was discovered [12] while the latter suggests that the outputs are cyclic or completely chaotic. Whichever the observed behavior, the recurrent network will produce a response (i.e., state vector) at each discrete-time step, which comprises the activation degree of all neurons of the model.

Although FCMs inherited many aspects from other neural models (i.e., the reasoning rule), there are some important differences regarding to other types of Artificial Neural Network (ANNs). Classical ANN models regularly perform like *black-boxes*, where both the neurons and the connections do not have a clear meaning for the problem itself, or results cannot easily be explained by the same predicting model. However, all neurons in an FCM have a precise meaning for the physical system being modeled and correspond to specific variables that form part of the solution. It should be highlighted that an FCM does not comprise hidden neurons since these entities could not be interpreted nor help at explaining why a solution is suitable for a given problem. If this were the case, the model becomes unfriendly for many further phases.

In the last years, FCMs have been widely studied due to its advantageous characteristics for handling complex systems. Less attention has been given to the development of FCM-based classifiers. *Pattern classification* [4] is one of the most ubiquitous real-world problems and certainly one at which humans really excel. It consists of identifying the right category (among those in a predefined set) to which an observed pattern belongs. These patterns are often described by a set of predictive attributes of numerical and/or nominal nature called features. Some successful classifiers include: artificial neural networks [7], support vector machines [8] or random forest [2]. Regrettably none of these black-box classifiers provides an inherent introspection into the reasoning process associated to the decision model. However, in some areas where machine learning models are applied, the transparency in their predictions is crucial.

Aiming at developing a novel classification model, Papakostas et al. [31, 32] introduced the notion of *FCM-based classifier*. The most prominent challenge to be confronted when constructing an FCM-based classifier relies on the approach to connect input and output neurons. It should be remarked that the topology of an FCM-based classifier must comprise a coherent and precise meaning for the physical system under investigation. This suggests that the intervention of human experts to define the network topology is usually required.

The development of accurate learning algorithms for computing the required parameters is another issue that deserves attention. In the literature, several unsupervised and supervised learning methods have been recently proposed [29]. These algorithms are mostly focused on computing the weight matrix that define the semantic of the whole cognitive system. However, the prediction capability of an FCM-based classifier does not exclusively depend on the weight set. Other aspects such as the network's capability the represent the problem domain or the convergence issues are equally important.

In this chapter, we focus on main advances on FCM-based classification and challenges that remain open problems for the scientific community. The rest of the manuscript is structured as follows. Section 2 briefly surveys theoretical aspects related to FCMs. Section 3 discusses about the transparency and usability of models for understanding the decision process. Section 4 describes the use of FCMs in the context of pattern classification. Section 5 describes the FCM-based models where input neurons denote information granules rather low-level features. To conclude, Sects. 6 and 7 will wrap-up the paper and highlight the main points of view of this proposal.

2 Fuzzy Cognitive Maps

FCMs can be seen as recurrent neural networks with interpretability features that have been widely used in modeling tasks [11]. They consist of a set of neural processing entities called concepts (neurons) and their causal relations. The activation value of such neurons regularly takes values in the [0, 1] interval, so the stronger the activation value of a neuron, the greater its impact on the network. Of course, connected weights are also relevant in this scheme. The strength of the causal relation between two neurons C_i and C_j is quantified by a numerical weight $w_{ij} \in [-1, 1]$ and denoted via a causal edge from C_i to C_j.

There are three possible types of causal relationships between neural processing units in an FCM-based network that express the type of influence from one neuron to the other, which are detailed as follows:

- If $w_{ij} > 0$ then an increase (decrement) in the cause C_i produces an increment (decrement) of the effect C_j with intensity $|w_{ij}|$.
- If $w_{ij} < 0$ then an increase (decrement) in the cause C_i produces an decrement (increment) of the neuron C_j with intensity $|w_{ij}|$.
- If $w_{ij} = 0$ then there is no causal relation between C_i and C_j.

Equation 1 shows the Kosko's activation rule, with $A^{(0)}$ being the initial state. This rule is iteratively repeated until a stopping condition is met. A new activation vector is calculated at each step t and after a fixed number of iterations, the FCM will be at one of the following states: (i) equilibrium point, (ii) limited cycle or (iii) chaotic behavior [12]. The FCM is said to have converged if it reaches a fixed-point attractor, otherwise the updating process terminates after a maximum number of iterations T is reached.

$$A_i^{(t+1)} = f\left(\sum_{j=1}^{M} w_{ji} A_j^{(t)}\right), i \neq j \tag{1}$$

The function $f(\cdot)$ in Eq. 1 denotes a monotonically non-decreasing nonlinear function used to clamp the activation value of each neuron to the allowed interval. Examples of such functions are the bivalent function, the trivalent function, and the sigmoid variants [37].

We put emphasis in the sigmoid function since it has exhibited superior prediction capabilities [3]. Equation 2 formalizes the non-linear transfer function used in our conducted researches, where λ is the sigmoid slope and h denotes the offset. Several studies reported at [1, 10, 14, 17, 27] have shown that such parameters are closely related with the network convergence.

$$f(A_i) = \frac{1}{1 + e^{-\lambda(A_i - h)}} \tag{2}$$

Equation 1 shows an inference rule widely used in many FCM-based applications, but it is not the only one. Stylios and Groumpos [36] proposed a modified inference rule, found at Eq. 3, where neurons take into account its own past value. This rule is preferred when updating the activation value of neurons that are not influenced by other neural processing entities.

$$A_i^{(t+1)} = f\left(\sum_{j=1}^{M} w_{ji} A_j^{(t)} + A_i^{(t)}\right), i \neq j \tag{3}$$

Another modified updating rule was proposed in [28] to avoid the conflicts emerging in the case of non-active neurons. Being more explicit, the rescaled inference depicted in Eq. 4 allows dealing with the scenarios where there is not information about an initial neuron-state and helps preventing the saturation problem. The reader can notice that we can obtain a similar effect by using a shifted sigmoid function with the adequate slope.

$$A_i^{(t+1)} = f\left(\sum_{j=1}^{M} w_{ji}(2A_j^{(t)} - 1) + (2A_i^{(t)} - 1)\right), i \neq j \tag{4}$$

If the cognitive network is able to converge, then the system will produce the same output towards the end, and therefore the activation degree of neurons will remain unaltered (or subject to infinitesimal changes). On the other hand, a cyclic FCM produces dissimilar responses with the exception of a few states that are periodically produced. The last possible scenario is related to chaotic configurations in which the network yields different state vectors. Formally, such situations are mathematically defined as follows:

- **Fixed-point** ($\exists t_\alpha \in \{1, 2, \ldots, (T-1)\} : A^{(t+1)} = A^{(t)}, \forall t \geq t_\alpha$): the FCM produces the same state vector after the t_α-th iteration-step. This suggests that $A^{(t_\alpha)} = A^{(t_\alpha+1)} = A^{(t_\alpha+2)} = \cdots = A^{(t)}$.
- **Limit cycle** ($\exists t_\alpha, P \in \{1, 2, \ldots, (T-1)\} : A^{(t+P)} = A^{(t)}, \forall t \geq t_\alpha$): the FCM periodically produces the same state vector after the t_α-th iteration-step. This suggests that $A^{(t_\alpha)} = A^{(t_\alpha+P)} = A^{(t_\alpha+2P)} = \cdots = A^{(t_\alpha+jP)}$ where $t_\alpha + jP \leq T$, such that $j \in \{1, 2, \ldots, (T-1)\}$.
- **Chaos**: the FCM continues producing different state vectors for successive cycles, thus being impossible to make suitable decisions.

If the FCM is unable to converge, then the model will produce confusing responses and thus a pattern cannot be derived [26], thus being impossible to arrive at suitable conclusions. In presence of chaos or cyclic situations, the reasoning rule stops once a maximal number of iterations T is reached. In classification scenarios, the decision class is then calculated from the last cycle, but this output is partially unreliable due to the lack of convergence.

3 The Reasoning Process and Its Explainability

The *classification problem* [4] is about building a mapping $f : \mathcal{U} \rightarrow \mathcal{D}$ that assigns to each instance $x \in \mathcal{U}$ described by the attribute set $\Phi = \{\phi_1, \ldots, \phi_M\}$ a decision class D from the N possible ones in $\mathcal{D} = \{D_1, \ldots, D_N\}$. The mapping is often learned in a *supervised* fashion, i.e., by relying on an existing set of previously labeled examples used to train the model. The learning process is regularly driven by the minimization of a cost/error function.

Researchers in *Machine Learning* are primarily focused on prediction rates. Regrettably, most accurate classifiers do not provide any mechanism to explain how they arrived at a particular conclusion and therefore behave like a “black-box”. Some classifiers like *Artificial Neural Networks*, *Support Vector Machines*, *Ensemble techniques* or *Random Forests* are well-known to be the most likely successful algorithms for addressing classification problems in terms of prediction rates. However, their lack of transparency negatively effects their usability in scenarios where understanding the decision process is required.

For example, neural computation is a widely studied research field within Artificial Intelligence. The main limitation of Artificial Neural Networks is their lack of transparency, which means that the network cannot justify its complex reasoning process. As a result, these models do not allow interpreting the semantic behind the physical system under investigation since the transparency is a necessary condition to build interpretable classifiers.

Aiming at elucidating the hidden reasoning process of black-boxes, several post-hoc procedures have been proposed. For example, one of these explanatory techniques used explicit IF-THEN rules for extracting knowledge from black-box classifiers while more recent procedures use symbolic approaches to approximate the model [9]. But whether such explanation is truly comprehensive and meaningful in the case of large trees or rule sets is questionable.

The transparency inherent to FCMs and their underlying neural foundations have motivated researchers to build interpretable FCM-based classifiers. In these models, the interpretability may be achieved through causal relations between neural entities defining the modeled system. Regrettably, building *accurate*, *truly interpretable* FCM-based classifiers involves difficult challenges.

4 Low-Level FCM-Based Classifiers

As already mentioned, FCMs have been widely studied due to their appealing properties for handling complex and dynamic systems, but the development of FCM-based classifiers has received less attention.

One of the firsts attempt to use FCMs in the context of pattern classification was implemented in [31, 32]. In these references, the authors defined the notion of *FCM-based classifiers* and proposed some generic architectures. The most prominent challenge to be faced when constructing an FCM-based classifier lies on how to connect input and output neurons.

It should be remarked that an FCM classifier's topology (i.e., concepts and causal relations) must comprise a coherent and precise meaning for the physical system being modeled. If the input neurons represent features of the classification problem, then we are in presence of a *low-level cognitive classifier* where neural processing units can be categorized as shown below:

Definition 1 We say that a neural processing entity C_i is an *independent input neuron* if its activation value does not depend on the other input neurons.

Definition 2 We say that a neural processing entity C_i is a *dependent input neuron* if its activation value is influenced by other connected neurons.

Definition 3 We say that a neural processing entity C_i in an FCM-based classifier is an *output neuron* if we can predict a decision class from its final activation value, which only depends on the connected input neurons.

Typically, independent and dependent input neurons are used to activate the cognitive networks since they often denote problem features. Output neurons, on the other hand, are used to compute the decision class for an initial activation vector. In the case of independent input neurons, they can propagate their initial activation vector and they are not influenced by any other input neurons, therefore their activation values remain static. Notice that the expert must ideally determine the role of each neurons and the way that input neurons are connected each other. In spite of this fact, Papakostas et al. [30] investigated three generic architectures for mapping the decision classes:

- **Class-per-output architecture**. Each decision class is mapped to an output neurons. Therefore, the predicted decision class corresponds to the label of the output neuron having the highest activation value.
- **Single-output architecture**. Each decision class is enclosed into the activation space of a single output neuron.

1. *Using a clustering approach*. Each class is associated with a cluster center. In the testing phase, the center having the closest distance to the projected activation value is assigned to the input instance.
2. *Using a thresholding approach*. Each decision class is associated with a pair of thresholds. In the testing phase, the interval comprising the projected activation value is then assigned to the input instance.

In these architectures, the human intervention is required during the construction stage, and even so, the supervised learning methods will probably fail in producing *authentic causal relations* since they just fit the model to the existing data. Therefore, we are losing the interpretability features attached to the network, although the decision process remains transparent.

On the other hand, the absence of hidden neural entities in these recurrent neural networks may probably lead to poor prediction rates. Aiming at boosting the prediction capability of FCM-based classifier, in [32] the authors put forth two hybrid typologies. Figures 1 and 2 show these typologies that include a *black-box* classifier to improve the overall prediction rates.

In the first model, the black-box produces a confidence degree per decision class. Sequentially, this vector is used as initial configuration for the FCM model that corrects the responses produced by the black-box. In the second model, the input neurons are also connected to output ones, so the predictions computed by the black-box classifier can be understood as a bias.

These hybrid models completely destroy the transparency attached to the cognitive network. If this happens, then, there is no real reason to use FCMs in classification scenarios, instead we may adopt black-box models such as Support Vector Machines, Neural Networks or Random Forests.

Another key element towards designing a low-level FCM-based classifier relies on the learning algorithm. The chief objective behind FCM learning has been to

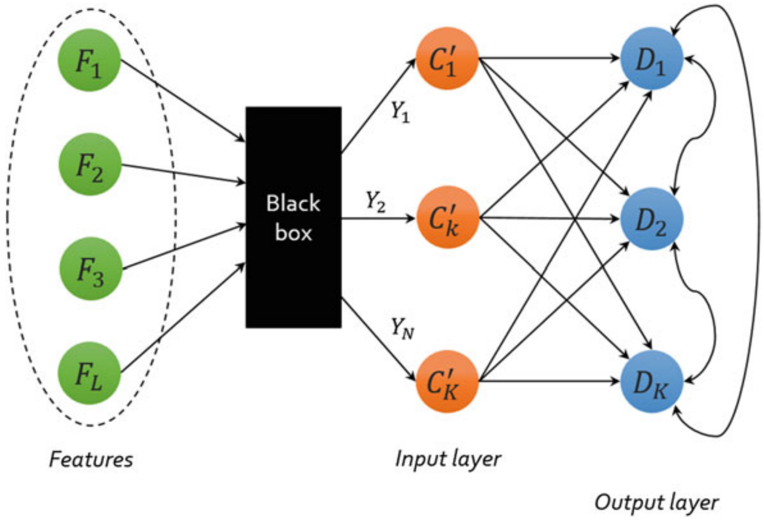


Fig. 1 Hybrid FCM-based classifier type-1

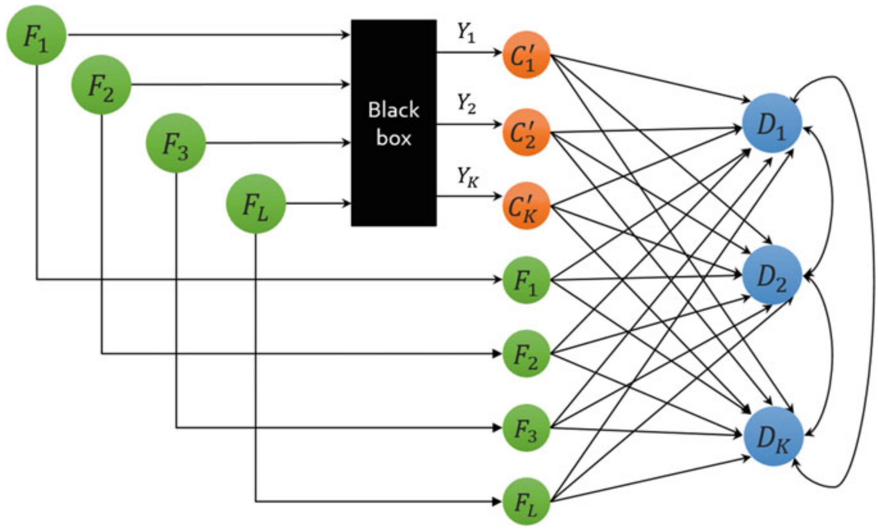


Fig. 2 Hybrid FCM-based classifier type-2

derive the weight matrix $W_{(M \times M)}$ that minimizes the prediction error based on expert intervention, available historical data or both. According to the their classification scheme, existing learning algorithms can be roughly gathered into two large groups: *unsupervised* and *supervised*.

4.1 Unsupervised Learning Algorithms

Hebbian-based learning methods are *unsupervised* procedures that do not require a set of labeled historical data, i.e., data in which the value of the decision feature(s) are previously known. The aim of learning FCMs by using adaptive Hebbian-based methods is to yield weight matrices on the basis of experts' knowledge and to improve the accuracy of previously set weights.

Papakostas et al. [30] thoroughly tested the performance of several Hebbian-type algorithms in classification scenarios, and concluded that these learning procedures regularly produce very poor classification rates.

More explicitly, Hebbian-type methods are convenient to fine-tune the weight set with a small deviation from the initial configuration. As a result, the adjusted causal relations partially preserve their physical meaning, which cannot be guaranteed when using a heuristic-based learning method. Of course, the requirement of experts' knowledge is a serious drawback. The flexibility on data requirements of these algorithms is the key aspect behind their poor generalization capability. This makes Hebbian-type algorithms unfit for solving pattern classification problems where multiple classes must be predicted.

4.2 Supervised Learning Algorithms

As an alternative to Hebbian-based methods, we can learn the network structure from data using heuristic-based algorithms [29] in a supervised fashion. Heuristic learning approaches aim at generating weight matrices that minimize an error function, viz. the difference between the expected responses and the map-inferred outputs. These methods are more expensive optimization techniques given that they regularly explore multiple candidate solutions. Besides, they require the definition of the objective function to be optimized, which is the core of these learning procedure, rather than the adopted search method.

Equation 5 formalizes an error function for pattern classification scenarios, where X denotes the weight matrix, K is the number of training instances, $\psi(.)$ is the decision model to be used for determining the class label, while S_k represents the expected decision class for the kth training instance. In the case of the single-output architecture, the class is computed from the activation value of the decision neuron at the last iteration-step.

$$E(X) = \frac{1}{k} \sum_{k=1}^{K} \begin{cases} \gamma_k, \ \psi(A_{Mk}^{(T)}) = S_k \\ 1, \ \psi(A_{Mk}^{(T)}) \neq S_k \end{cases} \tag{5}$$

Aiming at reducing the convergence error of the FCM-based classifier, the error function depicted in Eq. 5 uses a penalization factor γ_k for those instances that have been correctly classified. In short, the *convergence error* can be understood as the overall dissimilarity between the system response at each iteration, and the activation value at the last iteration-step.

Nápoles et al. [16, 17, 27] investigated the convergence of FCM-based classifiers and proposed a learning method to improve the system convergence, without altering the causal weights. More recently, they introduced an extended learning algorithm [26] where weights are estimated taking into account both accuracy and convergence. Based on these results, we propose a generalized measure to compute the convergence error of an FCM-based classifier.

Equation 6 shows the convergence error for the kth instance, where $\omega_t = t/T$ is the relevance of each iteration, M is the number of neurons, $N < M$ is the number of input-type ones, whereas $A_{ik}^{(t)}$ denotes the current activation value for the ith neuron. Moreover, π_k represents the centroid (ideal) point of the decision label associated to the kth training instance.

$$\gamma_k = \sum_{t=1}^{T} \frac{2\omega_t}{M(T+1)} \left(\sum_{i=1}^{N} \frac{(A_{ik}^{(t)} - A_{ik}^{(T)})^2}{N} + \sum_{i=1}^{M-N} \frac{(A_{ik}^{(t)} - \pi_k)^2}{M-N} \right) \tag{6}$$

Let us assume an FCM-based classifier using a single-output architecture and a thresholding approach, where the kth instance is associated with jth decision class. Equation 7 computes the centroid, where L_j^k and U_j^k denote the lower and upper decision thresholds, respectively.

$$\pi_k = \begin{cases} L_j^k, & L_j^k = 0 \\ U_j^k, & U_j^k = 1 \\ \frac{L_j^k + U_j^k}{2}, & L_j^k \neq 0, U_j^k \neq 1 \end{cases} \tag{7}$$

This approach introduces two key contributions in regard to the algorithm proposed in [26]. First, we remove the required parameters by measuring the convergence error if the target instance is correctly classified. This suggests that the system accuracy will always be favored. Second, we compute the converge error of sigmoid neurons according to their role in the network. The convergence error of input-type neurons is measured as the overall dissimilarity between the system response at each cycle, and the activation value at the last iteration. However, in the case of output-type neurons, we calculate the overall dissimilarity between each response, and the corresponding centroid.

Preliminary simulations using a Bioinformatic problem [21] have shown that this algorithms is capable of producing a suitable trade-off between convergence and accuracy. Ensuring the convergence helps in preventing the misclassifications of boundary instances, otherwise the model becomes fragile to perturbations. However, this algorithm cannot be generalized to other domains where the experts are unable to define the network topology.

5 High-Level Cognitive Classifiers

Cognitive mapping allows modeling different levels of interpretability, which depend on the abstraction degree. Neurons denoting entities with high abstraction level (i.e., information granules or prototypes) lead to high-level interpretable networks. If the level of abstraction is too high, then the physical system under investigation is difficult to analyze, so we are losing interpretability. On the other hand, defining attribute-level entities allow interpreting the system behavior at a low-level. However, sometimes the domain experts are unable to define precise, authentic causal relations with such specificity level.

High-level cognitive classifier refer to FCM-based models where input neurons denote information granules rather than low-level features. For example, Nápoles et al. [23, 24] introduced the notion of *rough cognitive mapping* in the context of pattern classification. The new classification model transforms the feature space into a granular one that is exploited using the neural inference rule present in FCM-based models. In these so-called *Rough Cognitive Networks* (RCNs), the weight matrix is automatically computed on the basis of the three-way decision rules [38] that construct three rough regions [33] to perform the classification process. The RCN model achieved competitive performance with respect to state-of-the-art methods in a real-world classification problems [23] as well as in a network intrusion detection scenario [22].

Figure 3 shows an RCN to solve any classification problem with two decision classes, where P_k, N_k and B_k are input neurons denoting the positive, negative and boundary regions related to the kth decision class.

More recently, two improved RCN models were introduced: *Rough Cognitive Ensembles* [20] and *Fuzzy-Rough Cognitive Networks* [25]. The purpose of these algorithms is to deal with the parametric requirement of rough cognitive classifiers while preserving their global prediction capabilities. The former is a granular ensemble model where each base RCN operates at a different granularity degree, whereas the latter replaced the crisp-rough constructs with fuzzy-rough ones. Numerical results have shown that both approaches are capable of outperforming the RCN algorithm. These modified algorithms perform comparably, thus we can achieve the same prediction rates using an ensemble composed of several networks that using a single fuzzy-rough classifier!

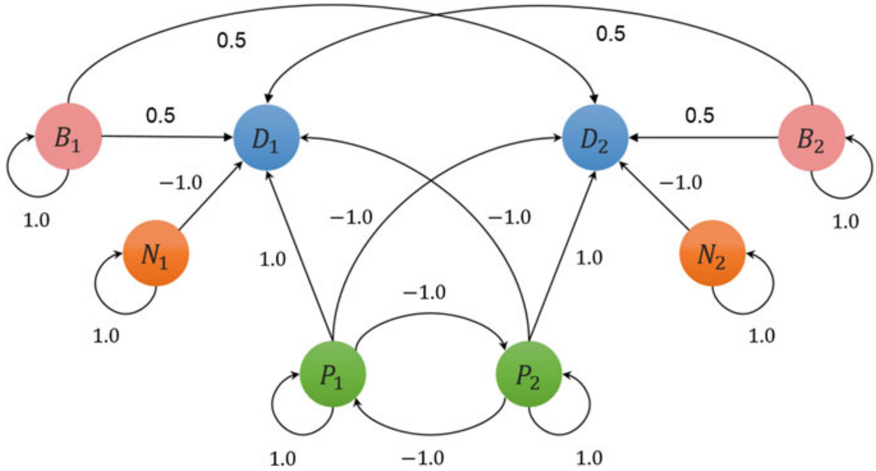


Fig. 3 RCN-based classifier for binary problems

Inspired on the RCN semantics and the approaches discussed in [34, 35], Nápoles et al. [19] proposed a *partitive granular cognitive map* to solve graded multi-label classification problems. In these machine learning problems, the goal is to predict the degree to which each instance relates to each available decision class. Three different FCM topologies were studied and several convergence features were included into the supervised learning methodology. Numerical experiments confirmed the ability of these granular classifiers to accurately estimate the degree of association between an object and each label.

It is worth highlighting the transparency on the decision model attached to Rough Cognitive Networks. In these models, we can interpret the physical system at a high-level by relying on the semantics behind the information constructs. However, a low-level reasoning is not possible, even when the classifier's decision process remains transparent and comprehensible.

6 Remaining Challenges

The development of accurate, truly interpretable FCM-based classifiers involves three main challenges, that still remain open:

- **Construction issues**. FCMs are knowledge-based techniques that regularly require the intervention of experts to define the network topology, i.e., the neurons and causal relations connecting them. Alternatively, we can learn the network structure from data using heuristic-based algorithms in a supervised fashion. However, these methods cannot produce *authentic causal relations* describing the system under analysis since they are oriented to fit the network to the historical data, without considering the system semantics. This implies that we cannot interpret the problem domain from such models, even if the FCM inference process is still transparent. Some authors attempt overcoming this drawback using correlation measures, which fail in capturing the underlying semantics behind causal relations. Being more explicit, it is well-known that causality does surely imply the existence of correlation, but the opposite does not necessarily hold.

- **Accuracy issues**. Generally speaking, the prediction rates of FCM-based classifiers are poor when compared with standard black-box models, mainly due to their limitations to represent the problem domain and the absence of theoretically sound learning algorithms. Papakostas et al. [30] concluded that Hebbian-based algorithms are not suitable in pattern classification environments, while the performance of heuristic-based learning methods quickly deteriorates when the number of neurons scales up. Froelich [5] proposed a promising post-optimization method to improve the prediction rates of FCM-based classifiers using a single-output architecture. Notice however that the overall prediction rates achieved by this method will heavily depend on the learning algorithm used to estimate the weight set.
- **Convergence issues**. FCM-based networks are recurrent cognitive systems that produce an output vector at each discrete-time step. This procedure is repeated until either the map converges or a maximal number of iterations is reached. Without ensuring the convergence, the model becomes unreliable and decision making becomes impossible. Regrettably, heuristic-based methods cannot ensure the FCM convergence, which implies that the resultant models are no longer interpretable and therefore, there is no reason to use cognitive mapping in pattern classification environments. More recently, Nápoles and his collaborators [16, 17, 26, 27] obtained promising results toward improving the convergence of FCM-based models without modifying the causal relations. However, analytical results reported in [18] have shown that establishing a suitable balance between convergence and accuracy cannot always be achieved without altering the weights.

It should be observed that the accuracy and convergence issues are mathematical challenges that can be present in other Machine Learning approaches. After all, the main purpose of traditional classifiers is to achieve the best possible prediction rates. The construction issues are, however, more delicate. Defining authentic causal relations between neural entities is the key aspect towards designing truly interpretable FCM-based systems. Otherwise, the model will produce misleading results when performing WHAT-IF simulations.

As far as we know, there is no learning method able to discover authentic causal structures from historical records due to the lack of well-established statistical tests for measuring causality. Even some authors affirm that the term “causality” is a philosophic concept that cannot possibly be measured in a numerical way without performing controlled experiments.

7 Conclusions

The use of FCMs for modeling real-life problems by recreating virtual scenarios have been demonstrated and reported in literature. These knowledge-based networks have been used as a modeling tool to analyze the behavior of complex systems, where it is very difficult to describe the entire system by a precise mathematical model. Consequently, it is easier and more practical to represent the decision-making process in a graphical way.

This paper explored the development of FCM-based classifiers and focused on the wide research avenues it provides. In spite of the detected shortcomings and challenges, the transparency inherent to cognitive mapping keeps motivating researchers to build interpretable FCM-based classifiers. In these models, the interpretability is achieved through causal relations between neurons defining the system under analysis. FCM-based models also provide other set of attractive characteristics: they are able to discover hidden patterns, are flexible, dynamic, combinable and tunable from different perspectives.

The FCM-based modeling approach allows building the network in presence of incomplete, conflicting or subjective information. Moreover, the inherent neural features of cognitive mapping provide a promising research avenue towards improving their accuracy in prediction scenarios. This suggests that FCM-based models could be as efficient as black-box models while retaining their ability to elucidate the system behavior through causal relations. Precisely, this conjecture, among other factors, keeps this research subject as a challenge open to the scientific community and a lively field of research.

Acknowledgements This work was partially supported by the Research Council of Hasselt University. Moreover, Gonzalo Nápoles would like to thank Frank Vanhoenshoven, Hasselt University, for his critical remarks and valuable suggestions.

References

1. Boutalis, Y., Kottas, T.L., Christodoulou, M.: Adaptive estimation of fuzzy cognitive maps with proven stability and parameter convergence. IEEE Trans. Fuzzy Syst. **17**(4), 874–889 (2009)
2. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
3. Bueno, S., Salmeron, J.L.: Benchmarking main activation functions in fuzzy cognitive maps. Expert Syst. Appl. **36**(3), 5221–5229 (2009)
4. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley (2012)
5. Froelich, W.: Towards improving the efficiency of the fuzzy cognitive map classifier. Neurocomputing **232**, 83–93 (2017)
6. Grau, I., Nápoles, G., Bonet, I., Garcia, M.M.: Backpropagation through time algorithm for training recurrent neural networks using variable length instances. Computación y Sistemas **17**(1), 15–24 (2013)
7. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall PTR, Upper Saddle River (1998)

8. Hearst, M.A., Dumais, S.T., Osman, E., Platt, J., Scholkopf, B.: Support vector machines. IEEE Intell. Syst. Appl. **13**(4), 18–28 (1998)
9. Jacobsson, H.: Rule extraction from recurrent neural networks: a taxonomy and review. Neural Comput. **17**(6), 1223–1263 (2005)
10. Knight, C.J., Lloyd, D.J., Penn, A.S.: Linear and sigmoidal fuzzy cognitive maps: an analysis of fixed points. Appl. Soft Comput. **15**, 193–202 (2014)
11. Kosko, B.: Fuzzy cognitive maps. Int. J. Man-Mach. Stud. **24**(1), 65–75 (1986)
12. Kosko, B.: Hidden patterns in combined and adaptive knowledge networks. Int. J. Approx. Reason. **2**(4), 377–393 (1988)
13. Kosko, B.: Fuzzy Engineering. Prentice Hall (1997)
14. Kottas, T.L., Boutalis, Y.S., Christodoulou, M.A.: Fuzzy cognitive networks: adaptive network estimation and control paradigms. In: Glykas, M. (ed.) Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools and Applications, pp. 89–134. Springer, Berlin, Heidelberg (2010)
15. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. In: Anderson, J.A., Rosenfeld, E. (eds.) Neurocomputing: Foundations of Research, pp. 15–27. MIT Press, Cambridge (1988)
16. Nápoles, G., Bello, R., Vanhoof, K.: Learning Stability Features on Sigmoid Fuzzy Cognitive Maps through a Swarm Intelligence Approach, pp. 270–277. Springer, Berlin, Heidelberg (2013)
17. Nápoles, G., Bello, R., Vanhoof, K.: How to improve the convergence on sigmoid fuzzy cognitive maps? Intell. Data Anal. **18**(6S), S77–S88 (2014)
18. Nápoles, G., Concepción, L., Falcon, R., Bello, R., Vanhoof, K.: On the accuracy-convergence trade-off in sigmoid fuzzy cognitive maps. IEEE Trans. Fuzzy Syst. (submitted) (2017)
19. Nápoles, G., Falcon, R., Papageorgiou, E., Bello, R., Vanhoof, K.: Partitive granular cognitive maps to graded multilabel classification. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1363–1370 (2016)
20. Nápoles, G., Falcon, R., Papageorgiou, E., Bello, R., Vanhoof, K.: Rough cognitive ensembles. Int. J. Approx. Reason. **85**, 79–96 (2017)
21. Nápoles, G., Grau, I., Bello, R., Grau, R.: Two-steps learning of fuzzy cognitive maps for prediction and knowledge discovery on the HIV-1 drug resistance. Expert Syst. Appl. **41**(3), 821–830 (2014)
22. Nápoles, G., Grau, I., Falcon, R., Bello, R., Vanhoof, K.: A granular intrusion detection system using rough cognitive networks. In: Abielmona, R., Falcon, R., Zincir-Heywood, N., Abbass, H. (eds.) Recent Advances in Computational Intelligence in Defense and Security, chapter 7. Springer (2016)
23. Nápoles, G., Grau, I., Papageorgiou, E., Bello, R., Vanhoof, K.: Rough cognitive networks. Knowl.-Based Syst. **91**, 46–61 (2016)
24. Nápoles, G., Grau, I., Vanhoof, K., Bello, R.: Hybrid model based on rough sets theory and fuzzy cognitive maps for decision-making. In: International Conference on Rough Sets and Intelligent Systems Paradigms, pp. 169–178. Springer (2014)
25. Nápoles, G., Mosquera, C., Falcon, R., Grau, I., Bello, R., Vanhoof, K.: Fuzzy-rough cognitive networks. Neural Netw. (2017)
26. Nápoles, G., Papageorgiou, E., Bello, R., Vanhoof, K.: Learning and convergence of fuzzy cognitive maps used in pattern recognition. Neural Process. Lett. 1–14 (2016)
27. Nápoles, G., Papageorgiou, E., Bello, R., Vanhoof, K.: On the convergence of sigmoid fuzzy cognitive maps. Inf. Sci. **349–350**, 154–171 (2016)
28. Papageorgiou, E.I.: A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. Appl. Soft Comput. **11**(1), 500–513 (2011)
29. Papageorgiou, E.I.: Learning algorithms for fuzzy cognitive maps—a review study. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **42**(2), 150–163 (2012)
30. Papakostas, G., Koulouriotis, D., Polydoros, A., Tourassis, V.: Towards hebbian learning of fuzzy cognitive maps in pattern classification problems. Expert Syst. Appl. **39**(12), 10620–10629 (2012)

31. Papakostas, G.A., Boutalis, Y.S., Koulouriotis, D.E., Mertzios, B.G.: Fuzzy cognitive maps for pattern recognition applications. Int. J. Pattern Recogn. Artif. Intell. **22**(8), 1461–1486 (2008)
32. Papakostas, G.A., Koulouriotis, D.E.: Classifying patterns using fuzzy cognitive maps. In: Glykas, M. (ed.) Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools and Applications, pp. 291–306. Springer, Berlin, Heidelberg (2010)
33. Pawlak, Z.: Rough sets. Int. J. Comput. Inf. Sci. **11**(5), 341–356 (1982)
34. Pedrycz, W.: The design of cognitive maps: a study in synergy of granular computing and evolutionary optimization. Expert Syst. Appl. **37**(10), 7288–7294 (2010)
35. Pedrycz, W., Homenda, W.: From fuzzy cognitive maps to granular cognitive maps. IEEE Trans. Fuzzy Syst. **22**(4), 859–869 (2014)
36. Stylios, C.D., Groumpos, P.P.: Modeling complex systems using fuzzy cognitive maps. IEEE Trans. Syst. Man Cybern.—Part A: Syst. Hum. **34**(1), 155–162 (2004)
37. Tsadiras, A.K.: Comparing the inference capabilities of binary, trivalent and sigmoid fuzzy cognitive maps. Inf. Sci. **178**(20), 3880–3894 (2008)
38. Yao, Y.: Three-way decisions with probabilistic rough sets. Inf. Sci. **180**(3), 341–353 (2010)

A Proposal of On-Line Detection of New Faults and Automatic Learning in Fault Diagnosis

Adrián Rodríguez Ramos, Alberto Prieto Moreno, Antônio José da Silva Neto and Orestes Llanes-Santiago

Abstract In this paper a new approach of automatic learning for a fault diagnosis system using fuzzy clustering techniques is presented. The proposal presents an off-line learning stage, for training the classifier to diagnose the initial known faults and the normal operation state. In this stage, the data are firstly pre-processed to eliminate outliers and reducing the confusion in the classification process by using the Density Objective Fuzzy C-Means (DOFCM) algorithm. Later on, the Kernel Fuzzy C-Means (KFCM) algorithm is used to achieve greater separability among the classes, and reducing the classification errors. Finally, a step is developed to optimize the two parameters used in the algorithms in the training stage using the Differential Evolution algorithm. After the training, the classifier is used on-line (recognition stage) in order to process every new sample taken from the process. In this stage, a novel fault detection algorithm is applied. The algorithm analyzes the observations which are not classified in the known classes and belonging to a window of time to determine if they constitute a new class, probably representative of a new fault or if they are noise. If a new class is identified, a procedure is developed to incorporate it to the known classes by the classifier. The approach proposed was validated using an illustrative example. The results obtained indicate the feasibility of the proposal.

A. Rodríguez Ramos · A. Prieto Moreno · O. Llanes-Santiago (✉)
Departamento de Automática y Computación, Universidad Tecnológica de la Habana José Antonio Echeverría, CUJAE, CUJAE Calle 114, No. 11901, 10390 La Habana, Cuba
e-mail: orestes@tesla.cujae.edu.cu

A. Rodríguez Ramos
e-mail: adrian.rr@automatica.cujae.edu.cu

A. Prieto Moreno
e-mail: albprieto@automatica.cujae.edu.cu

A.J. da Silva Neto
Instituto Politécnico da Universidade Do Estado Do Rio de Janeiro (IPRJ/UERJ), Nova Friburgo, Brazil
e-mail: ajsneto@iprj.uerj.br

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_6

1 Introduction

In current industries, there is a marked necessity to improve the processes efficiency in order to produce with higher quality besides attending the environmental and industrial safety regulations [18, 39]. In the industries, the faults in equipments can have an unfavorable impact in the availability of the systems, the environment and the safety of operators. For such reasons, the faults need to be detected and isolated, being these tasks associated to the fault diagnosis systems [34].

Within the fault diagnosis methods there are those based on models [7, 8, 13, 32, 39, 40] and those based on the process historical data [2, 3, 15, 31, 37]. In the first approach, the tools use models that describe the functioning of the processes. These tools are based on the residue generation obtained from the difference between the measurable signals from the real process and the values obtained from the model. This entails an elevated knowledge about the characteristics of the processes, their parameters, and operation zones. However, it is usually very difficult to achieve due to the complexity of the industrial processes. On the other hand, the approaches based in historical data do not need a mathematical model, and they do not require much prior knowledge of the process parameters [41]. These characteristics constitute an advantage for complex systems, where relationships among variables are nonlinear, not totally known, and therefore, it is very difficult to obtain an analytical model that describes efficiently the dynamics of the process.

The fault diagnosis systems based on historical data are trained to be able to classify the process states known by the experts. However, with the decrease of the useful life of the automation technical devices such as sensors, actuators and pumps among others, the probability of occurrence of new faults increases. In this situation, the diagnostic systems will not be able to correctly classify the new faults, which will cause an erroneous decision making. For this reason, the topic related with the automatic identification of new patterns has gained a great importance in the area of fault diagnosis, [12, 17], where the fault diagnosis systems are needed to detect the new faults and incorporate them in their knowledge base. In this way, the fault diagnosis systems will have an automatic learning mechanism to update their knowledge base.

By performing an analysis of the different techniques developed in the recent years for control and fault diagnosis tasks, it is significative the increment in the use of the fuzzy clustering methods [1, 5, 19, 33, 36].

Fuzzy clustering techniques are very important tools of unsupervised data classification [16], that can be used to organize data into groups based on similarities among the individual data. Fuzzy clustering deals with the uncertainty and vagueness that can be found in a wide variety of applications, such as: image processing, pattern recognition, object recognition, modeling and identification [20, 23, 25, 35, 38, 42, 44]. The main focus of all fuzzy clustering techniques is to improve the clustering by avoiding the influence of the noise and outlier data.

The Fuzzy C-Means (FCM) algorithm [4], is one of the most widely used algorithm for clustering due to its robust results for overlapped data. Unlike k-means algorithm, data points in the FCM algorithm may belong to more than one cluster.

FCM algorithm obtains very good results with noise free data but are highly sensitive to noisy data and outliers [16].

Other similar techniques such as, Possibilistic C-Means (PCM) [24] and Possibilistic Fuzzy C-Means (PFCM) [30] interpret clustering as a possibilistic partition and work better in presence of noise in comparison with FCM. However, PCM fails to find optimal clusters in the presence of noise [16] and PFCM does not yield satisfactory results when data set consists of two clusters which are highly unlike in size and outliers are present [16, 22].

Noise Clustering (NC) [10, 11], Credibility Fuzzy C-Means (CFCM) [9], and Density Oriented Fuzzy C-Means (DOFCM) [21] algorithms were proposed specifically to work efficiently with noisy data.

The clustering output depends upon various parameters such as distribution of data points inside and outside the cluster, shape of the cluster and linear or non-linear separability. The effectiveness of the clustering method strongly relies on the choice of the metric distance adopted. FCM uses Euclidean distance as the distance measure, and therefore, it can only be able to detect hyper spherical clusters. Researchers have proposed other distance measures such as, Mahalanobis distance measure, and Kernel based distance measure in data space and in high dimensional feature space, such that non-hyper spherical/non-linear clusters can be detected [45, 46].

Another problem usually present in fuzzy clustering methods is that their performance depend significantly on the initialization of their parameters. In many occasions, it is necessary to make multiple runs of the algorithm in order to obtain good results which is time consuming, and not always the obtention of the best solution is guaranteed.

In order to overcome these problems in this paper a new approach to automatic learning and on-line detection of new faults using fuzzy clustering techniques is proposed. The methodology presents an off-line training stage and an on-line recognition stage. In the training stage the historical data of the process are used to train a fuzzy classifier and the center of each one of known classes are calculated. In a first step, the data are pre-processed to eliminate outliers and reducing the confusion in the classification process. To accomplish this objective, the Density Objective Fuzzy C-Means (DOFCM) algorithm is used. Later on, the Kernel Fuzzy C-Means (KFCM) algorithm is used to achieve greater separability among the classes, and reducing the classification errors. Finally, an step is used to optimize the parameters m (factor that regulates the fuzziness of the resulting partition) and σ (bandwidth and indicates the degree of smoothness of the Gaussian kernel function) of the algorithms used in this stage, applying DE algorithm. After the training, the classifier is used on-line (recognition) in order to process every new sample taken from the process. In this stage, a new fault detection algorithm is applied. The algorithm analyzes the observations which are not classified in the known classes and belonging to a window of time to determine if they constitute a new class, probably representative of a new fault or if they are noise. If a new class is identified, a procedure is developed to incorporate it to the knowledge base of the classifier.

The principal contribution of this chapter is the obtaining of a methodology that adequately combines fuzzy clustering algorithms to solve the drawbacks of this type of techniques when the data is affected by noise and outliers, to improve the classification results by using kernel tools and to incorporate a new on-line fault detection algorithm as part of a mechanism of automatic learning.

The organization of the chapter is as follows: in Sect. 2 are presented the general characteristics of the tools used in the proposed methodology. In Sect. 3, a description of the new classification methodology using fuzzy clustering techniques is presented. The Sect. 4 presents an illustrative example used to validate the methodology proposed. Finally, the conclusions are presented.

2 General Description of the Computational Tools

In this section, a general description of the Density Oriented Fuzzy C-Mean algorithm and Kernel FCM is presented. In addition, the DE algorithm used to optimize the parameters of the DOFCM and KFCM algorithms in the training stage is described too.

2.1 Density Oriented Fuzzy C-Means (DOFCM)

The algorithm attempts to decrease the noise sensitivity in fuzzy clustering by identifying outliers before the clustering process. The DOFCM algorithm creates $c + 1$ clusters with c good clusters and one cluster of noise. This algorithm identifies outliers before the construction of the clusters, based on the density of data set, as is shown in Fig. 1.

The neighborhood of a given radius of each point in a data set has to contain at least a minimum number of other points. DOFCM defines a density factor, called the neighborhood membership, which express the measure density of an object in relation to its neighborhood. The neighborhood membership of a point i in X is defined as:

$$M^{i}_{neighborhood} = \frac{\eta^{i}_{neighborhood}}{\eta_{max}} \tag{1}$$

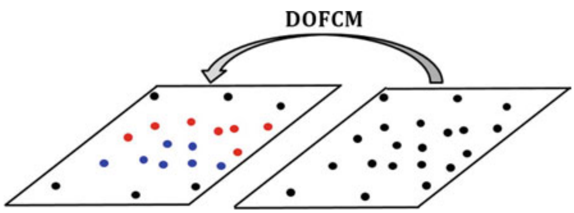


Fig. 1 Identification of outliers with the DOFCM algorithm

where $\eta^i_{neighborhood}$ is the number of points in the point neighborhood i; η_{max} is the maximum number of points in the neighborhood of any point in the data set.

If the point q is in the point neighborhood of the point i, q will satisfy:

$$q \in X | dist(i, q) \leq r_{neighborhood} \tag{2}$$

where $r_{neighborhood}$ is the radius of neighborhood, and $dist(i, q)$ is the distance between points i and q. Calculation of neighborhood radius is done as per [14]. Neighborhood membership of each point in the data set X is calculated using (1). The threshold value α is selected from the complete range of neighborhood membership values, depending on the density of points in the data set. The point will be considered as an outlier if its neighborhood membership is less than α. Let i be a point in the data set X, then

$$\begin{cases} M^i_{neighborhood} < \alpha \text{ then } i \;\; outlier \\ M^i_{neighborhood} \geq \alpha \text{ then } i \;\; non-outlier \end{cases} \tag{3}$$

α can be selected from the range of $M^i_{neighborhood}$ values after observing the density of points in the data set and it should be close to zero. Ideally, a point will be outlier only if no other point is present in its neighborhood i.e. when neighborhood membership is zero or threshold value $\alpha = 0$. However, in this scheme, a point is considered as an outlier when its neighborhood membership is less than α, where α is a critical parameter to identify outlier. Its value depends upon the nature of data set, i.e., taking into account the density of the data set, then, its value will vary for different data sets.

After identifying the outliers, the process of clustering begins. DOFCM reformulates FCM objective function as:

$$J_{DOFCM}(X; U, v) = \sum_{i=1}^{c+1} \sum_{k=1}^{N} \left(\mu_{ik}\right)^m \left(d_{ik}\right)^2 \tag{4}$$

where, the distances are defined by,

$$d^2_{ik} = \left(\mathbf{x}_k - \mathbf{v}_i\right)^T \mathbf{A}_i \left(\mathbf{x}_k - \mathbf{v}_i\right), \forall k, i = 1, \ldots, c \tag{5}$$

Membership function μ_{ik} is modified as:

$$\mu_{ik} = \begin{cases} \frac{1}{\sum_{j=1}^{c} \left(d_{ik}/d_{jk}\right)^{2/(m-1)}} & \text{if } non-outlier \\ 0 & \text{if } outlier \end{cases} \tag{6}$$

To update the centroid, DOFCM algorithm uses Eq. (7) as FCM algorithm. For the constraint on fuzzy membership, DOFCM algorithm uses Eq. (8). The DOFCM algorithm is presented in Algorithm 1.

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} \left(\mu_{ik}^m \mathbf{x}_k \right)}{\sum_{k=1}^{N} \mu_{ik}^m} \tag{7}$$

$$0 \le \sum_{i=1}^{c} \mu_{ik} \le 1, k = 1, 2, \ldots, N \tag{8}$$

Algorithm 1 Density Oriented Fuzzy C-Means (DOFCM)

Input: data with outliers $\mathbf{X}$, c, $\epsilon > 0$, $m > 1$, α, number of iterations
Output: data without outliers $\mathbf{X_p}$, fuzzy partition $\mathbf{U}$, class centers $\mathbf{V}$.
Identification of the outliers:
Calculate neighborhood radius.
Calculate $\eta^i_{neighborhood}$ according to (2).
Select η_{max}.
Calculate $M^i_{neighborhood}$ according to (1).
With the value of α, identify outliers according to (3).
Clustering process:
Initialize $\mathbf{U}$ to random fuzzy partition.
for $l = 1$ to $l = Itr\_max$ **do**
    Update classes centers according to (7).
    Calculate the distance d_{ik} according to (5).
    Update $\mathbf{U}$ according to (6).
    Verify stopping criterion.
end for

2.2 *Kernel Fuzzy C-Means (KFCM)*

KFCM represents the kernel version of FCM. This algorithm uses a kernel function for mapping the data points from the input space to a high dimensional space, as it is shown in Fig. 2.

KFCM algorithm modifies the objective function of FCM using the mapping $\mathbf{\Phi}$ as:

$$J_{KFCM} = \sum_{i=1}^{c} \sum_{k=1}^{N} \left(\mu_{ik} \right)^m \left\| \mathbf{\Phi}(\mathbf{x_k}) - \mathbf{\Phi}(\mathbf{v_i}) \right\|^2 \tag{9}$$

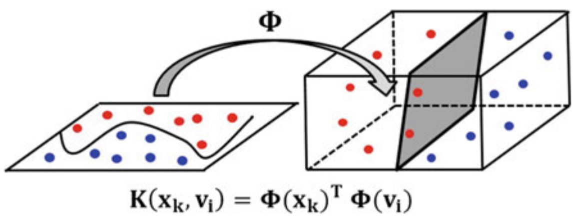


Fig. 2 KFCM feature space and kernel space

subject to:

$$\sum_{i=1}^{c} \mu_{ik} = 1, k = 1, 2, \ldots, N \tag{10}$$

where $\left\|\mathbf{\Phi}(\mathbf{x_k}) - \mathbf{\Phi}(\mathbf{v_i})\right\|^2$ is the square of the distance between $\mathbf{\Phi}(\mathbf{x_k})$ and $\mathbf{\Phi}(\mathbf{v_i})$. The distance in the feature space is calculated through the kernel in the input space as follows:

$$\begin{aligned}\left\|\mathbf{\Phi}(\mathbf{x_k}) - \mathbf{\Phi}(\mathbf{v_i})\right\|^2 = \mathbf{K}(\mathbf{x_k}, \mathbf{x_k}) - \mathbf{2K}(\mathbf{x_k}, \mathbf{v_i}) \\ + \mathbf{K}(\mathbf{v_i}, \mathbf{v_i})\end{aligned} \tag{11}$$

If the Gaussian kernel is used, then $\mathbf{K}(\mathbf{x}, \mathbf{x}) = \mathbf{1}$ and $\left\|\mathbf{\Phi}(\mathbf{x_k}) - \mathbf{\Phi}(\mathbf{v_i})\right\|^2 = \mathbf{2}\left(\mathbf{1} - \mathbf{K}(\mathbf{x_k}, \mathbf{v_i})\right)$. Thus (4) can be written as:

$$J_{KFCM} = 2 \sum_{i=1}^{c} \sum_{k=1}^{N} \left(\mu_{ik}\right)^m \left\|1 - \mathbf{K}(\mathbf{x_k}, \mathbf{v_i})\right\|^2 \tag{12}$$

where,

$$\mathbf{K}(\mathbf{x_k}, \mathbf{v_i}) = e^{-\left\|\mathbf{x}_k - \mathbf{v}_i\right\|^2/\sigma^2} \tag{13}$$

Minimizing (9) under the constraint shown in Eq. (10), yields:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{1 - \mathbf{K}(\mathbf{x_k}, \mathbf{v_i})}{1 - \mathbf{K}(\mathbf{x_k}, \mathbf{v_j})}\right)^{1/(m-1)}} \tag{14}$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} \left(\mu_{ik}^m \mathbf{K}(\mathbf{x_k}, \mathbf{v_i}) \mathbf{x_k}\right)}{\sum_{k=1}^{N} \mu_{ik}^m \mathbf{K}(\mathbf{x_k}, \mathbf{v_i})} \tag{15}$$

The KFCM algorithm is presented in Algorithm 2.

Algorithm 2 Kernel Fuzzy C-Means (KFCM)

Input: data $\mathbf{X_p}$, c, $\epsilon > 0$, $m > 1$, σ, number of iterations.
Output: fuzzy partition **U**, class centers **V**.
Initialize **U** to random fuzzy partition.
for $l = 1$ to $l = Itr\_max$ **do**
    Update classes centers according to (15) for Gaussian kernels.
    Calculate the distances according to (11).
    Update **U** according to (14).
    Verify stopping criterion.
end for

2.3 Differential Evolution (DE)

DE is an evolutionary algorithm based on populations, that uses methods derived from Biology like inheritance, mutation, natural selection and crossover. The idea behind DE is to generate a population of new feasible solutions based on perturbed solutions belonging to the previous population of solutions obtained up to a given time. This generation scheme is based on three operators: Mutation, Crossover and Selection. The configuration of DE can be summarized using the following notation:

$$DE/\mathbb{X}^j/\gamma/\lambda^*$$

where $\mathbb{X}^j$ denotes the solution to be perturbed in the j-th iteration, γ indicates the number of pairs of solutions to be used for perturbations of the current solution $\mathbb{X}^j$, and λ^* indicates the distribution function that will be used during the crossover. In the present work has been considered the configuration $DE/\mathbb{X}^{j(best)}/1/Bin$, where $\mathbb{X}^{j(best)}$ indicates the best individual of the population, Z, and Bin the Binomial Distribution function. This mutation operator is expressed in the following way:

$$\mathbb{X}^{j+1} = \mathbb{X}^{j(best)} + F_S(\mathbb{X}^{j(\alpha)} - \mathbb{X}^{j(\beta)}) \tag{16}$$

where $\mathbb{X}^{j+1}$, $\mathbb{X}^{j(best)}$, $\mathbb{X}^{j(\alpha)}$, $\mathbb{X}^{j(\beta)} \in \mathbb{R}^n$, $\mathbb{X}^{j(\alpha)}$ and $\mathbb{X}^{j(\beta)}$ are elements of the Z population and F_S is the escalation factor. For complementing the mutation operator, the crossover operator is defined for each component $\mathbb{X}_n$ of the solution vector:

$$\mathbb{X}_n^{j+1} = \begin{cases} \mathbb{X}_n^{j+1}, & \text{if } R < C_R \\ \mathbb{X}_n^{j(best)}, & \text{otherwise} \end{cases} \tag{17}$$

where $0 \le C_R \le 1$, is the crossover constant that is another control parameter in DE, and R is a random number generated by the distribution λ^* that in this case it is the binomial distribution.

Finally, the selection operator results as follows:

$$\mathbb{X}^{j+1} = \begin{cases} \mathbb{X}^{j+1}, & \text{if } F\left(\mathbb{X}^{j+1}\right) \le F\left(\mathbb{X}^{j(best)}\right) \\ \mathbb{X}^{j(best)}, & \text{otherwise} \end{cases} \tag{18}$$

The DE algorithm is presented in Algorithm 3.

Algorithm 3 Differential Evolution (DE)

Input: Z, F_S, C_R, number of iterations.
Output: $\mathbb{X}^{best}$ ($F(\hat{\theta})$=(m, σ))
Generate initial population of Z solutions.
Select better solution $\mathbb{X}^{best}$.
for $l = 1$ to $l = Itr\_max$ **do**
  Apply Mutation according to (16)
  Apply Crossover according to (17)
  Apply Selection according to (18)
  Update $\mathbb{X}^{best}$
  Verify Stopping criterion
end for

3 A New Classification Methodology with Automatic Learning Using Fuzzy Clustering

The new classification scheme with automatic learning proposed in this paper is shown in Fig. 3. It presents an off-line learning stage and an on-line recognition stage. In the training stage the historical data of the process are used to train (modeling the functional stages through the clusters) a fuzzy classifier. After the training, the classifier is used on-line (recognition) in order to process every new sample taken from the process. In this stage, the observations which are not classified in the known classes and belonging to a window of time are analyzed to determine if they

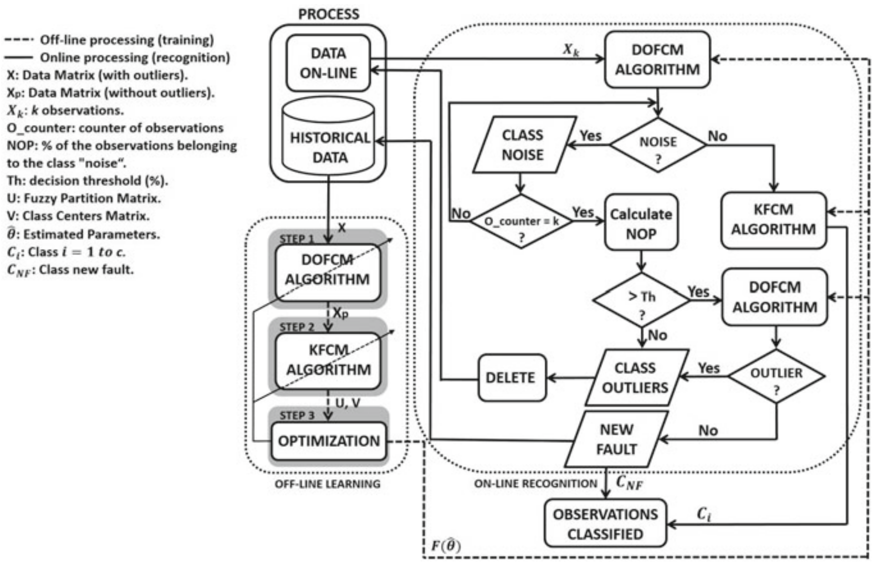


Fig. 3 Classification scheme using fuzzy clustering

constitute a new class, probably representative of a new fault or if they are noise. If a new class is identified, the experts characterize the new fault and it is incorporated to the knowledge base of the known faults and the classifier is trained again.

The clustering methods group the data in different classes based on a measure of similitude. In the processes, the data are acquired by means of a SCADA system (Supervisory Control and Data Acquisition), and the classes can be associated to functional states. In the case of statistical classifiers, each sample is compared with the center of each class by means of a measure of similitude to determine at which class the sample belongs. In the case of the fuzzy classifiers, the comparison is made to determine the membership degree of the sample to each class. In general, the higher membership degree determines the class which the sample is assigned, as it is showed in (19).

$$C_i = \left\{ i : max \left\{ \mu_{ik} \right\}, \forall i, k \right\} \tag{19}$$

3.1 Off-Line Training

In the first step, the center of the known classes $\mathbf{v} = \mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_N}$ is determined by using a historical data set representative of the different operation states of the process. In this step, a set of N observations (data points) $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ are classified into $c + 1$ groups or classes using the DOFCM algorithm. The c classes represent the normal operation conditions (NOC) of the process, and the faults to be diagnosed. They contain the information to be used in the next step. The other remaining class contains the data points identified as outliers by the DOFCM algorithm, and they are not used in the next step.

In the second step, the KFCM algorithm receives the set of observations classified by the DOFCM algorithm in the c classes. The KFCM algorithm maps these observations into a higher dimensional space in which the classification process obtains better results of satisfactory classifications. This step is shown in Algorithm (2). The Fig. 4 shows the procedure described in steps 1 and 2.

Finally, a step to optimize the parameters of the algorithms used in steps 1 and 2 is implemented. In this step, the parameters m and σ are estimated to optimize a validity index using an optimization algorithm. This will allow to obtain an improved U partition matrix, and therefore, a better position of the centers of the classes that characterize the different operation states of the system. Later, the estimated values of m in Eqs. (4), (12) and σ in Eq. (13) will be used during the on-line recognition, and it will contribute to improve the classification of the samples obtained by the data acquisition system from the system.

The validity measures are indexes to evaluate quantitatively the result of a clustering method and comparing its behavior when its parameters vary. Some indexes evaluate the resulting U matrix, while others are focused on the geometric resulting structure. The partition coefficient (PC) [26, 29, 43], which measures the fuzziness

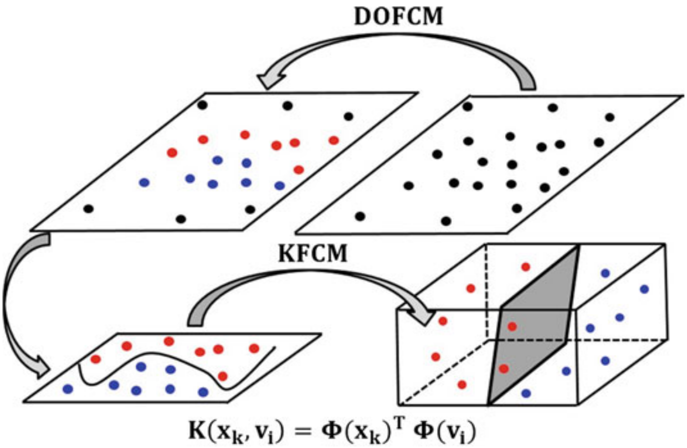


Fig. 4 Procedure performed by the DOFCM and KFCM algorithms

degree of the partition U, is used as validity measure in this case. Its expression is shown in the Eq. (20).

$$PC = \frac{1}{N} \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^2 \tag{20}$$

If the fuzziness degree of the partition U is high, the clustering process is better. Being analyzed in a different way, it allows to measure the degree of overlapping among the classes. In this case, the optimum comes up when PC is maximized, i.e., when each pattern belongs to only one group. Likewise, minimum comes up when each pattern belongs to each group.

Therefore, the optimization problem is defined as:

$$max\,\{PC\} = \frac{1}{N} \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^2$$

subject to:

$$m_{min} < m \leq m_{max}$$

$$\sigma_{min} \leq \sigma \leq \sigma_{max}$$

Then, a range of values of m and σ should be defined. It is known that $1 < m < \infty$, but from the practical point of view in many applications the value of m does not exceed of two [23, 25, 38, 44], therefore in this case: $1 < m \leq 2$. The parameter σ is called bandwidth, and it indicates the degree of smoothness of the function. In the case of σ if it is overestimated, the function tends to show a linear behavior and its projection in high-dimensional space loses its ability to separate non-linear data. Meanwhile, if σ is underestimated, the result will be highly sensitive to the noise presents in the data. Then, the search space of the algorithm must be large, so that during the exploration small and large values will be considered. In this chapter, a group of experiments were developed and it was found that an appropriate range was: $0.25 \leq \sigma \leq 20$.

In many scientific areas, and in particular in the fault diagnosis field, bio-inspired algorithms have been widely used with excellent results [6, 27, 28] to solve optimization problems. They can efficiently locate the neighborhood of the global optimum in most occasions with an acceptable computational time. There is a large number of bio-inspired algorithms, in their original and improved versions. Some examples are Genetic Algorithm (GA), Differential Evolution (DE), Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) among others. In this proposal, the DE algorithm, as described in Sect. 2.3 will be used to obtain the optimum values of the parameters m and σ due to its easy implementation and excellent outcomes.

3.2 On-Line Recognition

In this stage, a window of time with k observations and the parameter Th are established by the experts in this stage. The value of k is related with the characteristics of process. Th represents the percent of observations established by the experts to analyze if the observations classified as noise constitute a new class. If the observation is classified as a *good* sample the KFCM algorithm identifies to which of the known classes C_i belongs the observation. However, if the observation is classified as noise it is stored and a counter of noise observations (NO) is incremented. This procedure is repeated until the windows of time of k observations is completed.

After the k observations were classified, the percent of them classified as noise (NOP) is determined. If $NOP < Th$ the noise observations are not considered and the NO counter is restarted to begin a new cycle. If $NOP > Th$ the NO observations are analyzed to determine if their constitute a new class, probably representative of a new fault or they are outliers.

To analyze the noise observations, the DOFCM algorithm is used. It is based on the fact that the outliers are dispersed data with low density and do not form a cluster. However, when a new fault occurs the data will be concentrated (high density) by forming a cluster which characterize a new state.

The DOFCM algorithm is applied to the noise observations to determine based on the density of the data if they are outliers or represent the pattern of a new class. If the noise observations constitute a new class, the experts should identify whether the pattern corresponds to a single fault or the pattern is the result of several single faults by acting simultaneously (multiple fault). After identifying the pattern, it will be stored, if correspond, in the historical database used in the training stage. Later on, the classifier should be trained again and the procedure of online recognition will be repeated systematically.

The procedure explained for the on line stage represents a mechanism of online detection of novel faults with automatic learning for a fault diagnosis system. It is described in Algorithm 4.

Algorithm 4 Recognition

Input: data X_k, class centers $\mathbf{V}$, $r_{neighborhood}$, n_{max}, α, m, σ.
Output: Current State.
Select k
Select Th
Initialize $O\,counter = 0$
Initialize $NO\,counter = 0$
for $j = 1$ to $j = k$ **do**
  $O\,counter = O\,counter + 1$
  Calculate $\eta^i_{neighborhood}$ according to (2).
  Calculate $M^i_{neighborhood}$ according to (1).
  if $k \notin C_{outlier}$ **then**
    Calculate the distances from the observation k to class centers according to (11).
    Calculate the membership degree of the observation k to the c *good* classes according to (14).
    Determine to which class belongs the observation k using (19).
  else
    Store observation k in C_{noise}
    $NO\,counter = NO\,counter + 1$
  end if
end for
Calculate $NOP = \frac{NO\,counter}{O\,counter}$
if $NOP > Th$ **then**
  Apply DOFCM algorithm for C_{noise} considering only classes C_{NF} and $C_{outlier}$
  if $C_{noise} \notin C_{outlier}$ **then**
    Create a new fault: C_{NF}
    Store in the historical database.
  else
    Delete C_{noise}
    $NO\,counter = 0$
  end if
else
  Delete C_{noise}
  $NO\,counter = 0$
end if

---

4 Illustrative Examples with the Diamond Data Set

The diamond data set is used to present three illustrative examples of the proposed new fault diagnosis system with automatic learning using fuzzy clustering tools. This data set presents two classes with 5 observations each one, and each class has two variables: x_1 and x_2 [22, 30]. Figure 5 shows the diamond data set.

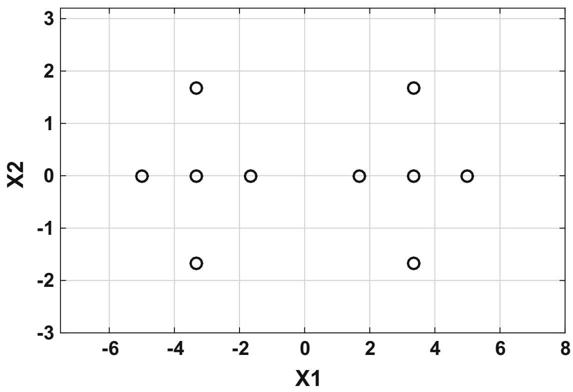


Fig. 5 Diamond data set

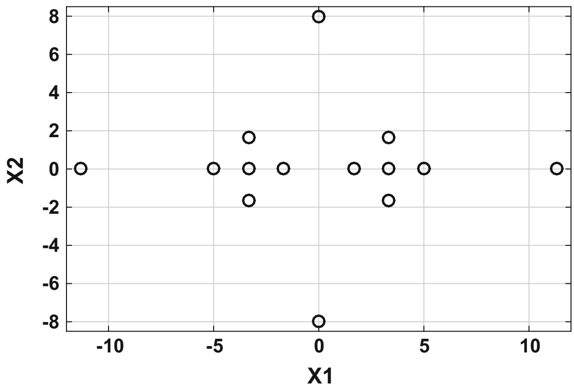


Fig. 6 Diamond data set modified for the training stage

4.1 Off-Line Training

Four new observations were added to the original data in order to represent the possible outliers. Figure 6 shows the diamond data set modified for the training stage.

To estimate the parameters mentioned in Sect. 3.1, *DE* algorithm was used due to its simple structure, and robustness [6]. The control parameters in *DE* are the size of the population *Z*, the crossover constant C_R and the scaling factor F_S. The values of the parameters for the DE algorithm considering a search space $1 < m \leq 2$ and $0.25 \leq \sigma \leq 20$ were: $C_R = 0.5$, $F_S = 0.1$ and $Z = 10$.

For the implementation of the *DE* algorithm the following stopping criteria are considered:

- Criterion 1: Maximum number of iterations.
- Criterion 2: Value of the objective function.

The behavior of the objective function (PC) is shown in Fig. 7 where it can be seen how its the value converges rapidly to one. Since the iteration number 4 the best parameters were obtained: $m = 1.0150$ and $\sigma = 15.3556$.

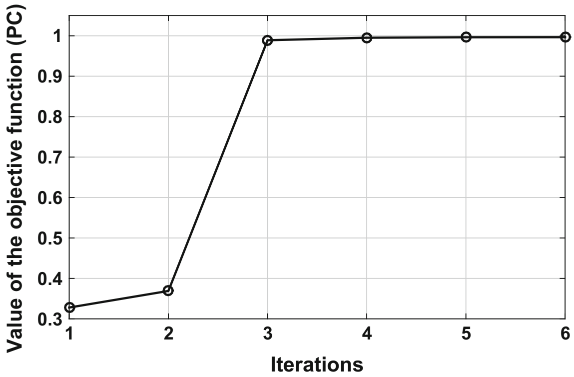


Fig. 7 Value of the objective function (PC)

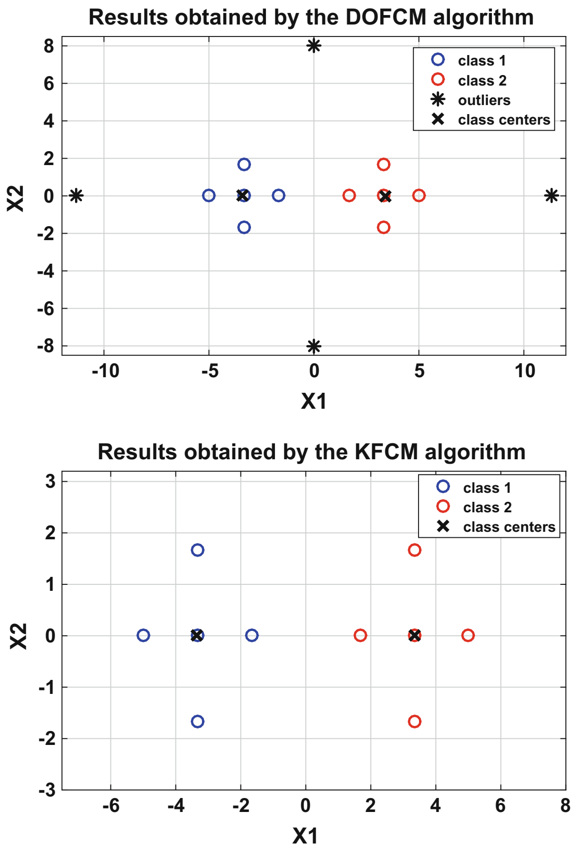


Fig. 8 Results of the training stage

Figure 8 shows the result of the clustering performed by the Algorithms 1 and 2 in the training stage. The values of the parameters used in these algorithms were: Number of iterations = 100, $\epsilon = 10^{-5}$ and $\alpha = 0.05$.

4.2 On-Line Recognition

Three examples were analyzed in order to test the recognition stage proposed in this chapter. Five observations were considered in all examples to simulate the observations sequentially obtained on-line. The numbers assigned to the added observations indicate the order of arrival of them. In the three examples, the value of the window of time of observations was $k = 5$, and a decision threshold of $Th = 50\%$ was used.

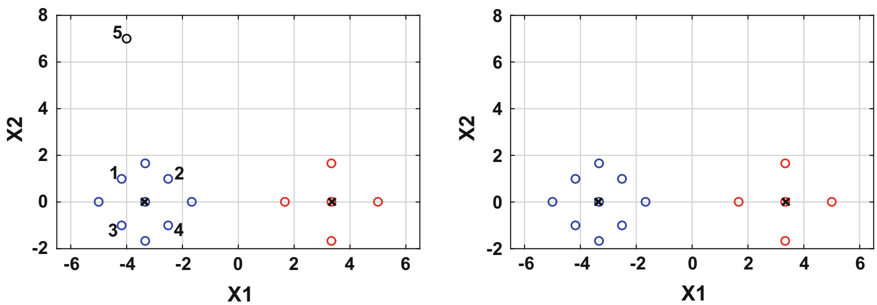
In Example 1, 5 observations are used, 4 belonging to class 1 and the other is an outlier. The objective is to evaluate the ability of a correct classification.

In Example 2, 2 observations belonging to class 2 and 3 outliers that did not form a class are used. The objective is to evaluate the correct classification and the ability to detect that the outliers do not conform a new class.

In Example 3, 5 observations belonging to a new class are used. The objective is to evaluate the correct classification of outliers and the ability to detect that these outliers conformed a new class.

Example 1 Four observations were classified in class 1 (in blue) and the other one was classified as outlier (in black) as it is shown in Fig. 9a. In this case $k = 5$, NO = 1, and NOP = 20%, then, the analyzes of the noise observations was not done. The final result of the on-line recognition is shown in Fig. 9b.

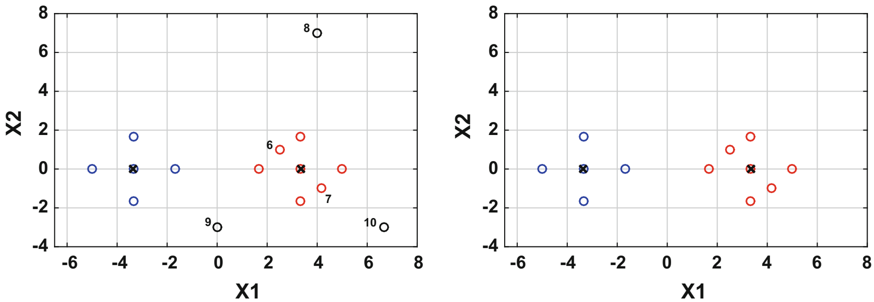
Example 2 Two observations were classified in class 2 (in red) and three observations were classified as outliers (in black) as is shown in Fig. 10a. In this case $k = 5$, NO = 3, and NOP = 60%, then, the analyzes of the noise observations was done. After the analysis of the three observations classified as outliers the diagnosis system decided that they did not represent a new class. The final result of the on-line recognition is shown in Fig. 10b.



(a) Results of the on-line classification after applying DOFCM algorithm in the example 1.

(b) Results after applying KFCM algorithm in the example 1.

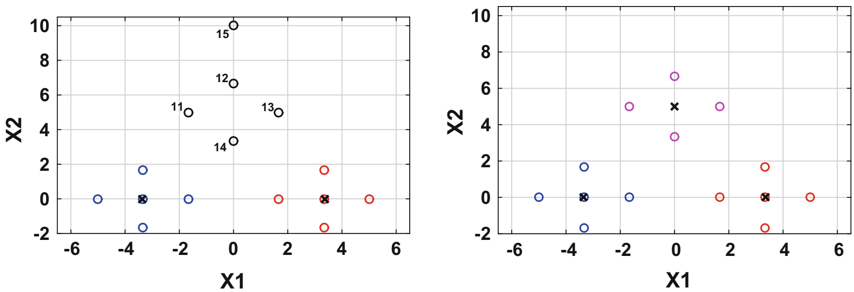
Fig. 9 Results of the recognition stage (Example 1)



(a) Results of the on-line classification after applying DOFCM algorithm in the example 2.

(b) Results after applying KFCM algorithm and to analyze the outliers in the example 2.

Fig. 10 Results of the recognition stage (Example 2)



(a) Results after applying DOFCM algorithm in the example 3

(b) Results after applying KFCM algorithm and to analyze the outliers in the example 3.

Fig. 11 Results of the recognition stage (Example 3)

Example 3 The five observations were classified as outliers (in black) as is shown in Fig. 11a. In this case $k = 5$, $\text{NO} = 5$, and $\text{NOP} = 100\%$, then, the analysis of the noise observations was done. After the analyzes four observations were classified in a new class (in purple) and the other was classified as outlier and it was not considered. The final result of the on-line recognition is shown in Fig. 11b.

5 Conclusions

In the present chapter a new classification scheme for on-line detection of new faults and automatic learning using fuzzy clustering techniques is proposed. In the proposal, the DOFCM algorithm is used in the first step of the training stage for preprocessing the data with the objective to remove the outliers. The KFCM algorithm is used in the second step of the training step for the data classification to make use of the advantages introduced by the kernel function in the separability of the classes, in order to obtain better classification outcomes. The algorithm DE is used to optimize the parameters of the DOFCM and KFCM algorithms. These parameters are used in the on-line recognition stage, where the classifier incorporates a new fault detection algorithm. In the on-line recognition stage, the proposed new algorithm analyzes the observations belonging to a window of time which are not classified in the known classes and determines if they constitute a new class, probably representative of a new fault or if they are noise. If a new class is identified, a procedure is developed to incorporate it to the base of knowledge of the classifier. The excellent results obtained show the feasibility of the proposal. For future research, an interesting idea is to design a fault diagnosis system with the ability to detect multiple new faults in the on-line recognition stage.

Acknowledgements The authors acknowledge the financial support provided by FAPERJ, Fundacão Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro; CNPq, Consehlo Nacional de Desenvolvimento Científico e Tecnológico; CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, research supporting agencies from Brazil; UERJ, Universidade do Estado do Rio de Janeiro and CUJAE, Universidad Tecnológica de La Habana José Antonio Echeverría.

References

1. Bedoya, C., Uribe, C., Isaza, C.: Unsupervised feature selection based on fuzzy clustering for fault detection of the tennessee eastman process. Adv. Artif. Intell. **7637**, 350–360 (2012). Springer LNAI
2. Bernal de Lázaro, J.M., Llanes Santiago, O., Prieto Moreno, A., Knupp, D.C., Silva-Neto, A.J.: Enhanced dynamic approach to improve the detection of small-magnitude faults. Chem. Eng. Sci. **146**, 166–179 (2016)
3. Bernal de Lázaro, J.M., Prieto Moreno, A., Llanes Santiago, O., Silva Neto, A.J.: Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems. Comput. Ind. Eng. **87**, 140–149 (2015)
4. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Springer, Plenum, New York (1981)
5. Botia, J., Isaza, C., Kempowsky, T., LeLann, M.V., Aguilar-Martín, J.: Automation based on fuzzy clustering methods for monitoring industrial processes. Eng. Appl. Artif. Intell. **26**, 1211–1220 (2013)
6. Camps Echevarría, L., Llanes Santiago, O., Silva Neto, A.J.: An approach for fault diagnosis based on bio-inspired strategies. Stud. Comput. Intell. **284**, 53–63 (2010)
7. Camps Echevarría, L., Llanes Santiago, O., Hernández Fajardo, J.A., Silva Neto, A.J., Jímenez Sánchez, D.: A variant of the particle swarm optimization for the improvement of fault diagnosis in industrial systems via faults estimation. Eng. Appl. Artif. Intell. **28**, 36–51 (2014)
8. Camps Echeverría, L., de Campos Velho, H.F., Becceneri, J.C., Silva Neto, A.J., Llanes Santiago, O.: The fault diagnosis inverse problem with ant colony optimization and ant colony optimization with dispersion. Appl. Math. Comput. **227**, 687–700 (2014)
9. Chintalapudi, K.K., Kam, M.: A noise resistant fuzzy c-means algorithm for clustering. In: IEEE Conference Fuzzy Systems Proceedings, vol. 2, pp. 1458–1463 (1998)

10. Dave, R.N.: Characterization and detection of noise in clustering. Pattern Recogn. Lett. **12**, 657–664 (1991)
11. Dave, R.N., Krishnapuram, R.: Robust clustering methods: a unified view. IEEE Trans. Fuzzy Syst. **5**, 270–293 (1997)
12. Detroja, K., Gudi, R., Patwardhan, S.: A possibilistic clustering approach to novel fault detection and isolation. J. Process Control **16**, 1055–1073 (2006)
13. Ding., S.X.: Model-Based Fault Diagnosis Techniques. Springer (2008)
14. Ester, M., Kriegel, H., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the 2nd ACM SIGKDD pp. 226–231, Portland, Oregon (1996)
15. Fan, J., Wang, Y.: Fault detection and diagnosis of non-linear non-gaussian dynamic processes using kernel dynamic independent component analysis. Inf. Sci. **259**, 369–379 (2014)
16. Gosain, A., Dahika, S.: Performance analysis of various fuzzy clustering algorithms: a review. In: 7th International Conference on Communication, Computing and Virtualization 2016, vol. 79, 100–111 (2016)
17. Hu, D., Sarosh, A., Dong, Y.F.: A novel kfcm based fault diagnosis method for unknown faults in satellite reaction wheels. ISA Trans. **51**, 309–316 (2012)
18. Hwang, I., Kim, S., Kim, Y., Seah, C.: A survey of fault detection, isolation, and reconfiguration methods. IEEE Trans. Control Syst. Technol. **18**, 636–656 (2010)
19. Jahromi, A.T., Er, M.J., Li, X., Lim, B.S.: Sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis. Neurocomputing **196**, 31–41 (2016)
20. Jiang, X.L., Wang, Q., He, B., Chen, S.J., Li, B.L.: Robust level set image segmentation algorithm using local correntropy-based fuzzy c-means clustering with spatial constraints. Neurocomputing **207**, 22–35 (2016)
21. Kaur, P.: A density oriented fuzzy c-means clustering algorithm for recognising original cluster shapes from noisy data. Int. J. Innovative Comput. Appl. **3**, 77–87 (2011)
22. Kaur, P., Soni, A., Gosain, A.: Robust kernelized approach to clustering by incorporating new distance measure. Eng. Appl. Artif. Intell. **26**, 833–847 (2013)
23. Kesemen, O., Tezel, O., Ozkul, E.: Fuzzy c-means clustering algorithm for directional data (fcm4dd). Expert Syst. Appl. **58**, 76–82 (2016)
24. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. IEEE Trans. Fuzzy Syst. **1**, 98–110 (1993)
25. Leski, J.M.: Fuzzy c-ordered-means clustering. Fuzzy Sets Syst. **286**, 114–133 (2016)
26. Li, C., Zhou, J., Kou, P., Xiao, J.: A novel chaotic particle swarm optimization based fuzzy clustering algorithm. Neurocomputing **83**, 98–109 (2012)
27. Liu, Q., Lv, W.: The study of fault diagnosis based on particle swarm optimization algorithm. Comput. Inf. Sci. **2**, 87–91 (2009)
28. Lobato, F., Steffen Jr., F., Silva Neto, A.: Solution of inverse radiative transfer problems in two-layer participating media with differential evolution. Inverse Prob. Sci. Eng. **18**, 183–195 (2009)
29. Pakhira, M., Bandyopadhyay, S., Maulik, S.: Validity index for crisp and fuzzy clusters. Pattern Recogn. **37**, 487–501 (2004)
30. Pal, N.R., Pal, K., Keller, J.M., Bezdek, J.C.: A possibilistic fuzzy c-means clustering algorithm. IEEE Trans. Fuzzy Syst. **13**, 517–530 (2005)
31. Pang, Y.Y., Zhu, H.P., Liu, F.M.: Fault diagnosis method based on kpca and selective neural network ensemble. Adv. Mater. Res. **915**, 1272–1276 (2014)
32. Patan, K.: Artificial Neural Networks for the Modelling and Fault Diagnosis of Technical Processes. Springer (2008)
33. Qin, S.J.: Survey on data-driven industrial process monitoring and diagnosis. Annu. Rev. Control **36**(2), 220–234 (2012)
34. Ramos, A.R., Acosta, C.D., Torres, P.R., Mercado, E.S., Baez, G.B., Rifón, L., Llanes-Santiago, O.: An approach to multiple fault diagnosis using fuzzy logic. J. Intell. Manuf. 1–11 (2016). doi:10.1007/s10845-016-1256-4

35. Saltos, R., Weber, R.: A rough-fuzzy approach for support vector clustering. Inf. Sci. **339**, 353–368 (2016)
36. Seera, M., Lim, C.P., Loo, C.K., Singh, H.: A modified fuzzy min-max neural network for data clustering and its application to power quality monitoring. Appl. Soft Comput. **28**, 19–29 (2015)
37. Sina, S., Sadough, Z.N., Khorasani, K.: Dynamic neural network-based fault diagnosis of gas turbine engines. Neurocomputing **125**, 153–165 (2014)
38. Thong, P.H., Son, L.H.: Picture fuzzy clustering for complex data. Eng. Appl. Artif. Intell. **56**, 121–130 (2016)
39. Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N.: A review of process fault detection and diagnosis, part 1: quantitative model-based methods. Comput. Chem. Eng. **27**, 293–311 (2003)
40. Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N.: A review of process fault detection and diagnosis, part 2: qualitative models and search strategies. Comput. Chem. Eng. **27**, 313–326 (2003)
41. Wang, J., Hu, H.: Vibration-based fault diagnosis of pump using fuzzy technique. Measurement **39**, 176–185 (2009)
42. Wong, K.I., Vonga, C.M., Wong, P.K.: Simultaneous-fault detection based on qualitative symptom descriptions for automotive engine diagnosis. Appl. Soft Comput. **22**, 238–248 (2014)
43. Wu, K., Yang, M.: A cluster validity index for fuzzy clustering. Pattern Recogn. **26**, 1275–1291 (2005)
44. Zhang, L., Lu, W., Liu, X., Pedrycz, W., Zhong, C.: Fuzzy c-means clustering of incomplete data based on probabilistic information granules of missing values. Knowl.-Based Syst. **99**, 51–70 (2016)
45. Zhang, D.Q., Chen, S.C.: Clustering incomplete data using kernel based fuzzy c-means algorithm. Neural Process Lett. **18**, 155–162 (2003)
46. Zhang, D.Q., Chen, S.C.: A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. Artif. Intell. Med. **32**, 37–50 (2004)

Fuzzy Portfolio Selection Models for Dealing with Investor's Preferences

Clara Calvo, Carlos Ivorra and Vicente Liern

Abstract This chapter provides an overview of the authors' previous work about dealing with investor's preferences in the portfolio selection problem. We propose a fuzzy model for dealing with the vagueness of investor preferences on the expected return and the assumed risk, and then we consider several modifications to include additional constraints and goals.

1 Introduction

H.M. Markowitz won the 1990 Nobel Price for his work in the foundation of modern portfolio theory (MPT) [27, 28], which has become a main tool in portfolio management as well as in other economic theories, such as asset pricing [33]. MPT is a deep theory which can hardly be described in a few words (see [11] for a comprehensive account), but, roughly speaking, it aims to determine the best portfolio we can form from a given set of possible assets on the basis of two characteristics. The first one is the expected return. In order to measure it, the return of each asset is considered as a random variable and the expected return is often measured by its mean, which in practice is estimated by the arithmetical mean of the historical returns. The expected return of a portfolio is defined as the weighted sum of the expected returns of its assets.

Here we should face the critical question: to what extent can we trust that the future return of a portfolio will be similar to the expected return calculated from its past returns? This leads to the second characteristic to be considered in order to select a portfolio to invest in: the risk. It tries to estimate the difference between the expected return and the real future return of a portfolio.

C. Calvo · C. Ivorra · V. Liern (✉)
Universidad de Valencia, Valencia, Spain
e-mail: vicente.liern@uv.es

C. Calvo
e-mail: clara.calvo@uv.es

C. Ivorra
e-mail: carlos.ivorra@uv.es

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_7

Whereas the theoretical relevance of MPT is not questioned, several criticisms about its real world applicability have arisen [38]. However there are also renowned specialists supporting it [18, 30, 33]. Assuming that the expected return and an adequate measure of the risk are reliable, MPT establishes that a rational investor should select an *efficient portfolio,* i.e., a portfolio providing the least possible risk for a given expected return or, what is essentially the same, providing the greatest expected return for a maximum allowable risk.

The original (classical) Markowitz model is also called the mean-variance model since it takes as indicators of the expected return and the risk of a portfolio the mean and the quadratic form associated to the variance-covariance matrix of the returns of the assets, respectively, which in practice are estimated from the historical data by standard statistical techniques assuming that they are normally distributed.

However, several alternative ways for measuring the risk of a portfolio have been proposed. Value at Risk (VaR) is widely used (see [36] for a discussion of this concept, or also [16]). Konno and Yamazaki [21] propose a linear model which dramatically simplifies the computational aspects of the portfolio selection problem. Some other possibilities arose from the fact that many risk measures become high when there is a high probability that the return will be far from the expected return, but they do not distinguish whether the difference is positive (higher return than expected) or negative (less return than expected). Since aversion to having more benefits than expected is questionable, some asymmetric measures for the so-called downside risk that take into account only the risk of having less return than expected, have been proposed. The first downside risk measure appeared in Roy's "safety-first" model [31, 32]. See also [8, 9, 25, 26, 35]. Other asymmetric measures of risk take into account higher statistical moments: skewness, kurtosis [15, 17, 20]. For more advanced models taking into account the dynamics of the variance see for instance [10].

2 The Classical Portfolio Selection Problem

Thus, the original Markowitz Portfolio Selection Model is formulated as

$$\begin{aligned} \text{Min. } & \mathbf{x^t V x} \\ \text{s.a } & \mathbf{e}^t\mathbf{x} \geq r \\ & \mathbf{1}^t\mathbf{x} = 1 \\ & \mathbf{x} \geq 0 \end{aligned} \tag{1}$$

where the vector $\mathbf{x}$ contains the weights of the assets in the portfolio (i.e. the proportion of each asset in the total invested fund), $\mathbf{e}$ is the vector of expected returns, measured by the means of the historical data and $\mathbf{V}$ is the variance-covariance matrix of such data, so that R estimates the risk of the portfolio and r is the minimum expected return specified by the investor.

Alternatively, a dual form of the problem consists of maximizing the expected return and imposing a maximum admissible risk:

$$\begin{aligned} \text{Max. } & \mathbf{e}^t\mathbf{x} \\ \text{s.a } & \mathbf{x}^t\mathbf{V}\mathbf{x} \leq R \\ & \mathbf{1}^t\mathbf{x} = 1 \\ & \mathbf{x} \geq 0 \end{aligned} \tag{2}$$

In fact, the portfolio selection problem is better understood as a bi-objective problem aiming both maximizing the expected return and minimizing the risk. The minimizing formulation is the most widely used in the literature, mainly because by being a quadratic problem it is more easily handled from a mathematical point of view. However, we will also deal with (2) since it is more realistic to ask an investor what risk he considers acceptable rather than forcing him to fix a minimum return without having any reference about the risk it carries. In fact, it is the usual practice for small investors [see for instance (http://www.santander.com)].

The original Markowitz Portfolio Selection Model included just linear constraints, mainly because computers could not handle more difficult instances. However, nowadays the available computational power is much greater and hence more sophisticated models can be dealt with, looking for efficient portfolios satisfying also additional constraints. There are many contexts in which such constraints become necessary. Some of them are related to the mutual fund management. Fund managers must comply contractual requirements determined by the prospectus as well as legal requirements, such as the 5–10–40-constraint imposed by the §60(1) of the German investment law [2], which establishes that securities of the same issuer are allowed to amount to up to 5% of the net asset value of the mutual fund, but they are allowed to 10% if the total of all of these assets is less than 40% of the net asset value. It is also usual to include buy-in thresholds to reduce transaction costs. This means not allowing the stocks of a mutual fund in a given asset to be less than a certain amount. A third typical example is that managers often impose upper bounds to the total number of assets in a mutual fund also to reduce transaction costs, as well as lower bounds in order to diversificate the investment. See [13] for the computational aspects associated to these additional constraints. This leads to the model

$$\begin{aligned} \text{Min. } & \mathbf{x}^t\mathbf{V}\mathbf{x} \\ \text{s.t. } & \mathbf{e}^t\mathbf{x} \geq r \\ & \mathbf{1}^t\mathbf{x} = 1 \\ & \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \end{aligned} \tag{3}$$

where $\mathbf{l}$ and $\mathbf{u}$ are the vectors of lower and upper bounds for each weight. However, in many cases the investor does not really wish to force each asset to have a minimum weight in the portfolio but, in order to avoid an artificially imposed excess of diversification in the optimal portfolio, he may just wish to require a minimum weight only for those assets actually appearing in it. This leads to the incorporation of semicontinuous variable constraints into the model, which means that each variable x_i is allowed either to be 0 or to vary in the rank $[l_i, u_i]$. Such constraints can be expressed with the help of binary variables y_i taking the value 1 if the i-th asset appears in the portfolio and 0 otherwise. The resulting model is:

$$\begin{aligned} \text{Min.}\, & R = \mathbf{x}^t \mathbf{V} \mathbf{x} \\ \text{s.t.}\quad & \mathbf{e}^t \mathbf{x} \geq r \\ & \mathbf{1}^t \mathbf{x} = 1 \\ & l_i y_i \leq x_i \leq u_i y_i,\ 1 \leq i \leq n, \\ & y_i \in \{0, 1\}. \end{aligned} \tag{4}$$

In any portfolio selection problem, the set of optimal pairs (r, R) consisting of the minimal risk R providing a given expected return r, or, equivalently, the maximum expected return provided by a portfolio that does not exceed a maximum level of risk R, is known as the efficient frontier of the problem. In the simplest case where even the sign constraints are removed, it consists just of a branch of parabola [6, 11]. However, in the last decades, computation techniques have been developed to solve large and more sophisticated instances of the portfolio selection problem including many different kinds of constraints, making it more realistic (see [12, 37]).

3 A Fuzzy Formulation of the Portfolio Selection Problem

Obviously, the portfolio selection, like most financial problems, is related with uncertainty because it consists of taking a decision about future events. Moreover, it is not easy to model the investor's preferences. After the seminal work by Markowitz, attention has been given to the study of alternative models [22, 25] which try to deal more efficiently with the uncertainty of the data. Most of these models are based on probability distributions, which are used to characterize risk and return. However, another way of dealing with uncertainty is to work with models based on soft computing. Watada [41] solves this problem by using imprecise aspiration levels for an expected biobjective approach, where the membership functions of the goals are of a logistic-type. In 2000, Tanaka et. al. [39] propose using possibility distributions to model uncertainty on the expected returns and to incorporate the knowledge of financial experts by means of a possibility degree of similarity between the future state of financial markets and the state in previous periods [14]. Multiobjective programming has also been used to design fuzzy models of portfolio selection, either for compromise solutions [29] or by introducing multi-indices [1]. Specific methods have even been proposed for dealing with the unfeasibility provoked by conflict between the expected return and the investor's diversification requirements [23, 24].

However, in this section we will consider a very different class of vagueness related to the portfolio problem, namely the vagueness of the invertor's criteria for selecting a satisfactory trade off between the risk he considers acceptable and the return he wishes to obtain. In other words, the investor must choose a point at the efficient frontier of the problem. From a theoretical point of view, the investor's preferences are usually formalized by means of utility functions, so that the final choice is that efficient portfolio maximizing a given utility function, but when we try to reflect the preferences of a real specific investor we must ask him directly for a point in the efficient frontier. Nevertheless, it is obvious that the investor's preferences are essentially vague, so that it is unnatural to force him to choose a specific point. In practice, he could only determine a zone or a fuzzy point on it.

This leads to the fuzzy model proposed by the authors in [5]. The main idea is to consider partially feasible solutions involving slightly greater risk than that fixed by the decision-maker, and to study the possibilities that they offer in order to improve the expected return.

A fuzzy set $\tilde{S}$, of partially feasible solutions, is defined so that the membership degree of a given portfolio depends on how much its risk exceeds the risk R_0 fixed by the investor. On the other hand, a second fuzzy set $\tilde{G}$ is defined, whose membership function reflects the improvement on the return provided by a partially feasible solution with respect to the optimal crisp return z^*. In practice, we consider piecewise linear membership functions

$$\mu_{\tilde{S}}(x) = \begin{cases} 1 & \text{if } r \le R_0, \\ 1 - \frac{r-R_0}{p_f} & \text{if } R_0 < r < R_0 + p_f, \\ 0 & \text{if } r \ge R_0 + p_f, \end{cases}$$

$$\mu_{\tilde{G}}(x) = \begin{cases} 0 & \text{if } z \le z^*, \\ \frac{z-z^*}{p_g} & \text{if } z^* < z < z^* + p_g, \\ 1 & \text{if } z \ge z^* + p_g, \end{cases}$$

where r and z are the risk and the return provided by the portfolio x (which is assumed to satisfy the constraints of (MV), except the second one); the parameter p_f is the maximum increment in the risk that the decision-maker can accept, and p_g is the increment on the return that the decision-maker would consider completely satisfactory. From this, we can define a global degree of satisfaction

$$\lambda(x) = \min\{\mu_{\tilde{G}}(x), \mu_{\tilde{S}}(x)\},$$

which is the membership degree for the fuzzy intersection of $\tilde{S} \cap \tilde{G}$. The fuzzy portfolio model becomes

Table 1 Returns on five assets

Year	AmT	ATT	USS	GM	ATS
1937	−0.305	−0.173	−0.318	−0.477	−0.457
1938	0.513	0.098	0.285	0.714	0.107
1940	0.055	0.2	−0.047	0.165	−0.424
1941	−0.126	0.03	0.104	−0.043	−0.189
1942	−0.003	0.067	−0.039	0.476	0.865
1943	0.428	0.3	0.149	0.225	0.313
1944	0.192	0.103	0.26	0.29	0.637
1945	0.446	0.216	0.419	0.216	0.373
1946	−0.088	−0.046	−0.078	−0.272	−0.037

$$\begin{aligned}(\text{FMV})\ \text{Max.}\ &\lambda(x)\\ \text{s.t.}\ &x \in \tilde{S}.\end{aligned} \tag{5}$$

In [3] exact and heuristic procedures for solving this problem are described. In order to illustrate the main idea on which the model is based we consider five assets from the historical data introduced by Markowitz [28]. Table 1 shows the returns of American Tobacco, AT&T, United States Steel, General Motors and Atcheson & Topeka & Santa Fe.

We have fixed a risk level $R = 0.03$. The optimal crisp portfolio is formed by assets AmT, ATT, GM, ATS and provides an optimal return $z^* = 0.103926$. For the fuzzy model, we have fixed $p_f = 0.02$, $p_g = 0.02$. By explicitly solving the Kuhn-Tucker conditions associated to the model, we can calculate the optimal return for a given risk R, which happens to be

$$F(R) = \frac{-0.02355 + 52.6832R + 2.6136\sqrt{-0.77841 + 52.6832R}}{9.09494 \times 10^{-13}R + 33.84051\sqrt{-0.77841 + 52.6832R}}$$

Computations are valid for risks in the interval $I = [0.025826, 0.083341]$. The functions $\mu_f(R)$ and $\mu_g(R)$ are shown in Fig. 1. They intersect at $R^* = 0.041381$, corresponding to $\lambda = 0.430977$. The return on the fuzzy portfolio is 0.112545, whereas the crisp return was 0.103926.

We observe that the global degree of satisfaction is low. This means that the risk is increased much more than the return of the asset. The higher value of λ, the more preferable the alternative fuzzy portfolio is. High fuzzy satisfaction levels are more usual when additional constraints are considered making the efficient frontier more irregular, as we will see in the next section.

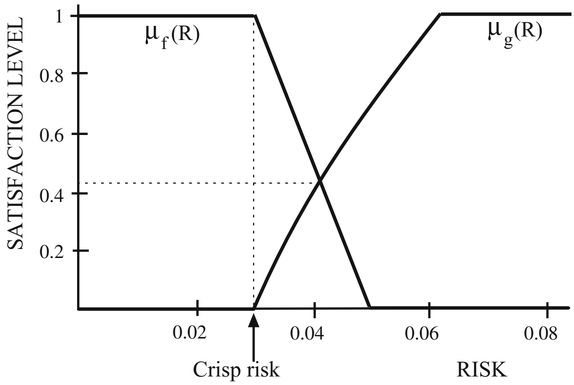


Fig. 1 Membership functions for the degree feasibility and degree of improvement on the expected return

4 Portfolio Selection with Semi-continuous Variable and Cardinality Constraints

As we have already noticed, real world investments require incorporating many additional constraints into the portfolio selection model, many of which can be expressed as mathematically simple linear constraints, but some others are more complex from a mathematical point of view since they transform the model into a mixed integer one. To illustrate this fact we will consider semicontinuous variable and cardinality constraints, although any set of linear constraints could be considered in addition. Hence, our starting point is now the model (4).

Let us call X the set defined by the constraints imposed on the problem when the minimum return constraint is relaxed, which will be handled separately. The constraint and goal set will be fuzzy subsets of the (crisp) universe set X. The fuzzy constraint set $\tilde{C}$ must be such that the value $\mu_{\tilde{C}}(\mathbf{x})$ is high when the expected return on the portfolio $\mathbf{x} \in X$ is not much less than r_0 and the risk is not much greater than R_0. This means that $\tilde{C}$ can be defined as the fuzzy intersection of two fuzzy sets $\tilde{C}_r$ and $\tilde{C}_R$, such that the degree of membership of each portfolio $\mathbf{x} \in X$ is given by $\mu_{\tilde{C}_r}(\mathbf{x}) := f_1(r(\mathbf{x}))$ and $\mu_{\tilde{C}_R}(\mathbf{x}) := g_1(R(\mathbf{x}))$, where $r(\mathbf{x})$ and $R(\mathbf{x})$ are the expected return and risk of the portfolio $\mathbf{x}$, $f_1 : \mathbb{R} \longrightarrow [0,1]$ is a non-decreasing function such that $f_1(r_0) = 1$ and $g_1 : \mathbb{R} \longrightarrow [0,1]$ is a non-increasing function such that $g_1(R_0) = 1$. The specific choice of f_1 and g_1 will depend on the available information about the investor's preferences regarding risk and return. Hence the membership function to the fuzzy feasible set $\tilde{C} := \tilde{C}_r \cap \tilde{C}_R$ is given by the membership function $\mu_{\tilde{C}}(\mathbf{x}) := \min\{\mu_{\tilde{C}_r}(\mathbf{x}), \mu_{\tilde{C}_R}(\mathbf{x})\}$, which is of the form $\mu_{\tilde{C}}(\mathbf{x}) = h_1(r(\mathbf{x}), R(\mathbf{x}))$, where $h_1(r, R) := \min\{f_1(r), g_1(R)\}$ generally has the shape shown in Fig. 2a.

On the other hand, the degree of membership of the goal set $\tilde{G}$ of the fuzzy problem must be high for portfolios whose expected return is much greater than r_0 or the risk is much less than R_0. Hence, $\tilde{G}$ is the fuzzy union of the fuzzy sets $\tilde{G}_r$ and $\tilde{G}_R$ whose membership functions are of the form $\mu_{\tilde{G}_r}(\mathbf{x}) := f_2(r(\mathbf{x}))$ and $\mu_{\tilde{G}_R}(\mathbf{x}) := g_2(R(\mathbf{x}))$, where $f_2 : \mathbb{R} \longrightarrow [0,1]$ is a non-decreasing function such that

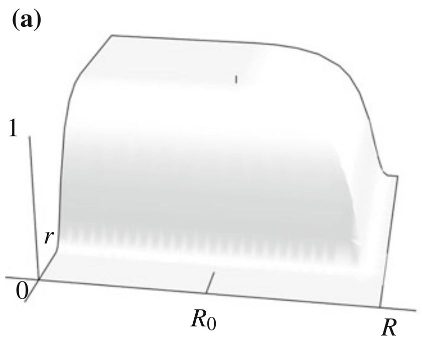
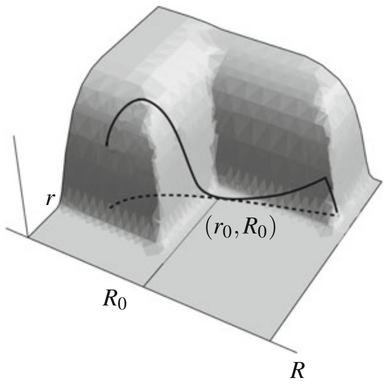


Fig. 2 Membership functions to $\tilde{C}$ and $\tilde{G}$

Fig. 3 Degree of global satisfaction and the efficient frontier. The surface represents the degree of global satisfaction for each pair (R, r) of expected risk and desired return. The *dotted line* is the efficient frontier and the *continuous line* is its lifting to the graph



$f_2(r_0) = 0$ and $g_2 : \mathbb{R} \longrightarrow [0, 1]$ is a non-increasing function such that $g_2(R_0) = 0$. Notice that in this case $\tilde{G}$ is a fuzzy union and not a fuzzy intersection, since improving the crisp optimal portfolio means improving the risk *or* improving the expected return, but both cases cannot occur simultaneously. The fuzzy intersection would be the empty set.

Thus, the membership function of the fuzzy goal set $\tilde{G} = \tilde{G}_r \cup \tilde{G}_R$ (which can be called the *degree of improvement of the goal*) is given by

$$\mu_{\tilde{G}}(\mathbf{x}) := \max\{\mu_{\tilde{G}_r}(\mathbf{x}), \mu_{\tilde{G}_R}(\mathbf{x})\} = h_2(r(\mathbf{x}), R(\mathbf{x})),$$

where $h_2(r, R) := \max\{f_2(r), g_2(R)\}$ has the shape shown in Fig. 2b.

Now, following Delgado et al. [7], we consider the fuzzy decision set of our problem, defined as the fuzzy intersection $\tilde{D} := \tilde{C} \cap \tilde{G}$, which has the shape shown in Fig. 3. The degree of membership of a portfolio $\mathbf{x}$ to $\tilde{D}$ is called its degree of global satisfaction: $\lambda(\mathbf{x}) := \min\{\mu_{\tilde{C}}(\mathbf{x}), \mu_{\tilde{G}}(\mathbf{x})\}$. In Fig. 3 we have represented a possible (simplified) efficient frontier of the crisp problem (SCP) and the pair (r_0, R_0) chosen by the investor as the starting point for the fuzzy model. We see that its degree of feasibility is 1 but its degree of improvement of the goal is 0, and so the degree of global satisfaction is 0. We can also see the lifting of the efficient frontier to the graph of the degree of global satisfaction. In order to choose a specific solution from the fuzzy decision set $\tilde{D}$ we maximize its degree of global satisfaction, i.e. we find the optimal fuzzy portfolio by solving the program:

$$\begin{aligned} \text{(FSC) Max. } & \lambda(\mathbf{x}) \\ \text{s.t. } & \mathbf{x} \in X \end{aligned}$$

In Fig. 3 we can see that the degree of global satisfaction has two local maxima on the efficient frontier, the best of which is the optimal solution of the fuzzy model we are introducing.

The problem of choosing membership functions suitable for modelling a real uncertain situation is a very subtle issue in fuzzy set theory. Here, in the absence of specific preferences, we will consider the simplest case. Notice that we intend to compare possible variations of the expected return with possible variations of the risk, and what is really comparable with a variation of the expected return is not a variation of its variance but a variation of its typical deviation. The difference between the variance and the typical deviation is just a square root, which is irrelevant when minimizing the risk, but it must be incorporated into our membership functions. In other words, the natural way to express the investor's preferences on the trade-off between variations in the expected return and variations in the risk is in terms of the mean and the typical deviation instead of the mean and the variance. In the absence of more specific criteria, we will assume a piecewise linear dependence on r and $\sqrt{R}$; namely, we take

$$f_1(r) := \begin{cases} 0 & \text{if } r < r_0 - p_{f_1}, \\ 1 - \frac{r_0 - r}{p_{f_1}} & \text{if } r_0 - p_{f_1} \le r \le r_0, \\ 1 & \text{if } r_0 < r, \end{cases} \quad g_1(R) := \begin{cases} 1 & \text{if } \sqrt{R} < \sqrt{R_0}, \\ 1 - \frac{\sqrt{R} - \sqrt{R_0}}{p_{g_1}} & \text{if } \sqrt{R_0} \le \sqrt{R} \le \sqrt{R_0} + p_{g_1}, \\ 0 & \text{if } \sqrt{R} > \sqrt{R_0} + p_{g_1}, \end{cases}$$

$$f_2(r) := \begin{cases} 0 & \text{if } r < r_0, \\ \frac{r - r_0}{p_{f_2}} & \text{if } r_0 \le r \le r_0 + p_{f_2}, \\ 1 & \text{if } r_0 + p_{f_2} < r, \end{cases} \quad g_2(R) := \begin{cases} 1 & \text{if } \sqrt{R} < \sqrt{R_0} - p_{g_2}, \\ \frac{\sqrt{R_0} - \sqrt{R}}{p_{g_2}} & \text{if } \sqrt{R_0} - p_{g_2} \le \sqrt{R} \le \sqrt{R_0}, \\ 0 & \text{if } \sqrt{R_0} < \sqrt{R}. \end{cases}$$

In [5] we show how to handle this fuzzy model. As an illustration, we consider the same data set considered in the previous section, but now we incorporate semicontinuous variables with vectors of bounds $\mathbf{l} = (0.2, 0.3, 0.2, 0.3, 0.2)$ and $\mathbf{u} = (0.6, 0.6, 0.6, 0.6, 0.6)$, as well as a cardinality constraint with $m = 2$ and $M = 5$.

Let us consider an investor that has chosen an expected return $r_0 = 0.125$, whose corresponding risk is $R_0 = 0.0742$. In order to interpret this variance, we will calculate the standard deviation $\sqrt{R_0} = 0.272$. We can consider the later as a quite high risk, and so we assume that the investor would be interested in reducing it. In this sense, a reduction of $p_{g_2} = 0.06$ would be considered as totally satisfactory. On the other hand, an increment greater than $p_{g_1} = 0.01$ would not be acceptable in any case. We assume that the investor would accept variations on the expected return with tolerances $p_{f_1} = p_{f_2} = 0.02$.

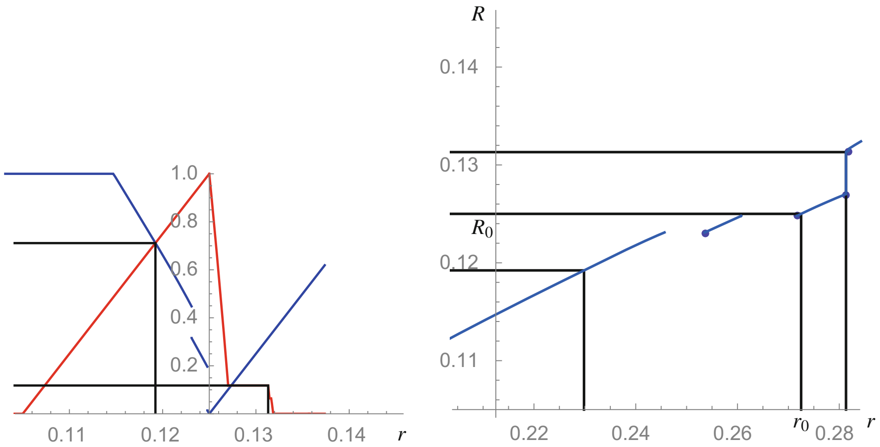


Fig. 4 *Left* Membership function to the feasible set and to the goal set as functions of the expected return. *Right* The efficient frontier around the crisp expected return

Table 2 Comparison between the crisp and the fuzzy solutions of Example 1

	x_1	x_2	x_3	x_4	x_5	r	$\sqrt{R}$
Crisp	0.31486	0	0.2	0.486	0	0.125	0.273
FzL	0.244	0.354	0	0.40	0	0.1192	0.23
FzR	0.5	0	0	0.3	0.2	0.1313	0.281

	λ	Δr	ΔR
Fuzzy left	0.71	-0.0058	-0.013
Fuzzy right	0.12	0.0063	0.008

Figure 4 left shows the degrees of membership to the feasible and the goal sets as functions of the expected return on a given efficient portfolio. We can compare it with the piece of the efficient frontier around r_0 within the tolerance levels, which is shown at the right.

There we can see that near r_0 there are two horizontal jumps below and a vertical one above. In Fig. 4 we have highlighted the two local maxima of the degree of global satisfaction. Specifically, they correspond to the efficient portfolios described in Table 2, which also includes the crisp efficient portfolio. Both in the figure and in the table we can see that the fuzzy optimal solution is the left-hand one with a degree of global satisfaction $\lambda = 0.71$. Notice that the three portfolios shown in Table 2 have different compositions.

In Table 2 we can also appreciate the interest of the fuzzy alternative: by changing from the crisp portfolio to the fuzzy one, we reduce the risk of the investment by a bit more than 1% at the cost of reducing the expected return by just 0.0058. The significantly lower degree of global satisfaction of the right fuzzy solution is reasonable since the increment on the expected return is far less than the increment on the risk.

5 Portfolio Selection with Non-financial Goals

The Social Investment Forum in its new 2012 Trends Report in US [40] finds that 11.23% of all assets under professional management in the United States at the end of 2011 applied various environmental social, governance and ethical criteria in their investment analysis. Investors practicing Socially Responsible Investment (SRI) strategies held \$3.74 trillion out of \$33.3 trillion of investment assets. This represents an increase of 22% since 2009 and reflects the "growing investors' interest in considering environmental, community, other societal or corporate governance (ESG) issues to refine how they make decisions as they select and manage their portfolios or raise their voices as shareholders" [40].

This growth of SRI strategies all around the world has stimulated in turn the rise of many entities working in the rating of assets with regard to their social responsibility. This poses two mathematical problems: how to evaluate assets' social responsibility which is by its nature a vague an imprecise concept and how to aggregate in a final rating a great amount of relevant but imprecise information about firms and/or funds.

Nevertheless, and although investors could be provided with highly processed non-financial information from the experts, in order to select a portfolio, they must elicit their preferences. The simplest way would be to restrict the feasible set of investments to those being "acceptable" for the investor from a SRI point of view. However this would mean to completely subordinate the financial goals to the non-financial ones and in fact, in practice most of the SRI assets first apply financial screens and then social screens. This clearly reflects that actually most of socially responsible investors consider SRI as a secondary goal with regard to maximizing the financial return and minimizing the financial risk.

In this section a fuzzy portfolio selection problem is proposed in which a secondary goal besides the financial ones is considered in such a way that no potentially interesting solution with regard to the risk and the return is discarded by the constraints. Specifically, the constraints of the model do not mention the secondary goal, which appears just in the objective function, in such a way that all the feasible portfolios within given ranges of risk and return are taken into consideration. Thus, the investor can be aware of what is being exactly missed as a result of the improvement of the additional non-financial goal.

Our starting point is again the model (4), and hence each portfolio is determined by a pair $(\mathbf{x}, \mathbf{y})$ of weights and binary variables. We measure the Social Responsibility of a portfolio as the degree of membership of a fuzzy set $\tilde{S}$, say $\mu_{\tilde{S}}(\mathbf{x}, \mathbf{y})$. See [4] for a way of defining such a fuzzy set.

Now we take as fuzzy feasible set the fuzzy subset of the set of all portfolios satisfying the hard constraints of (4) (i.e., all but the first one), defined as $\tilde{C} = \tilde{C}_r \cap \tilde{C}_R$, where the membership functions of the fuzzy sets $\tilde{C}_r$ and $\tilde{C}_R$ are given by:

$$\mu_{\tilde{C}_r}(x,y) = \begin{cases} 1 & \text{if } r \geq r_0, \\ \frac{r - r_0 + s_r}{s_r} & \text{if } r_0 - s_r < r < r_0, \\ 0 & \text{if } r \leq r_0 - s_r, \end{cases}$$

$$\mu_{\tilde{C}_R}(x,y) = \begin{cases} 1 & \text{if } R \leq R_0, \\ \frac{R_0 + s_R - R}{s_R} & \text{if } R_0 < R < R_0 + s_R, \\ 0 & \text{if } R \geq R_0 + s_R, \end{cases}$$

where r y R are respectively the expected return and the risk of the portfolio (x,y) and the values r_0, R_0, s_r and s_R are determined from the investor's preferences. This means that r_0 and R_0 are an expected return and a risk that the investor considers as completely acceptable, but he would accept worse values until reaching the tolerances s_r and s_R if this provides better results for the secondary goal.

Next we define a fuzzy goal set $\tilde{G}$ from two auxiliary fuzzy sets $\tilde{E}$ and $\tilde{S}$, the first one defining the "efficient enough" portfolios and the second one defining the "good enough" ones with regard to the secondary goal (always according to the investor's preferences). The set $\tilde{E}$ will express what we are loosing by accepting a non-efficient portfolio, and so efficient portfolios will be now the totally efficient portfolios, i.e. those having degree of membership of $\tilde{E}$ equal to 1.

First we define efficiency with regard to the expected return and then, the efficiency with regard to the risk by means of two fuzzy sets $\tilde{E}_r$ and $\tilde{E}_R$. The membership of $\tilde{E}_r$ is:

$$\mu_{\tilde{E}_r}(x,y) = \begin{cases} 1 - \frac{r_{\mathrm{ef}}(R) - r}{t_r} & \text{if } r \geq r_{ef}(R) - t_r, \\ 0 & \text{otherwise}, \end{cases}$$

where t_r is a tolerance determined from the investor's preferences and $r_{\mathrm{ef}}(R)$ is the efficient expected return corresponding to the risk R of the portfolio (x,y). This means that the degree of efficiency with regard to the expected return reaches the value 0 when the difference between the expected return r of the portfolio and $r_{\mathrm{ef}}(R)$ exceeds a tolerance fixed by the investor.

Analogously, we define the membership function of $\tilde{E}_R$ as

$$\mu_{\tilde{E}_R}(x,y) = \begin{cases} 1 - \frac{R - R_{\mathrm{ef}}(r)}{t_R} & \text{if } R \leq R_{ef}(r) + t_R, \\ 0 & \text{otherwise}, \end{cases}$$

which means that the degree of efficiency of a portfolio with regard to the risk is 1 for efficient portfolios and reaches the value 0 when the difference between the risk R of the portfolio and the efficient risk $R_{\mathrm{ef}}(r)$ for its return r exceeds a given tolerance t_R.

Now we define $\tilde{E} = \tilde{E}_r \cap \tilde{E}_R$, where the membership function of the fuzzy intersection is defined as the minimum of the previously defined membership functions. Hence the set $\tilde{E}$ allows us to speak about partially efficient portfolios in such a way that efficient portfolios in the usual sense are now the totally efficient ones, but a portfolio close enough to the efficient frontier is considered as "almost efficient" in the fuzzy sense.

Finally, we define our fuzzy goal set $\tilde{G}$ by means of the membership function as a weighted sum

$$\mu_{\tilde{G}}(x, y) = w\mu_{\tilde{S}}(x, y) + (1 - w)\mu_{\tilde{E}}(x, y),$$

where the weight w expresses the importance of the secondary goal for the investor with regard to efficiency. So, a high value for w means that the investor is willing to go relatively far from the efficient frontier in order to obtain higher values of $\mu_{\tilde{S}}$, whereas a small value of w means that the investor wishes to stay near the efficient frontier. In any case, recall we have defined the feasible set in such a way that only good enough solutions with regard to the financial goals are under consideration, and so the financial goals are always the main goals of the problem. More specifically, a large value for w means that, among the acceptable solutions with regard to the financial goals, those best with regard to $\tilde{S}$ are preferred, and only for similar values with regard to $\tilde{S}$ the degree of efficiency becomes relevant.

All in all, the degree of membership of the decision set is given by

$$\mu_{\tilde{D}}(x, y) = \min\{\mu_{\tilde{C}}(x, y), w\mu_{\tilde{S}}(x, y) + (1 - w)\mu_{\tilde{E}}(x, y)\}$$

and the fuzzy problem (6) is the problem determined by this decision set, whose optimal solutions are those with maximum degree of membership of $\tilde{D}$:

$$\begin{aligned} &\text{Max.}\min\{\mu_{\tilde{C}}(x, y), w\mu_{\tilde{S}}(x, y) + (1 - w)\mu_{\tilde{E}}(x, y)\} \\ &\quad \text{s.t.}\, 1x = 1 \\ &\qquad m \le \textstyle\sum_i y_i \le M \\ &\qquad l_i y_i \le x_i \le u_i y_i, \qquad i = 1, \dots, n \\ &\qquad x_i \ge 0,\ y_i \in \{0, 1\} \quad i = 1, \dots, n \end{aligned} \tag{6}$$

In order to illustrate this model, we consider the 10 mutual funds listed in Table 3. The first five have positive SRI degree, whereas the last five are conventional funds with null SRI degree. The variance-covariance matrix and the vector of expected returns are calculated from the weekly data from 31-12-2006 to 31-12-2007 provided by Morningstar Ltd. Assume we wish to select a portfolio consisting of a minimum of 3 and of a maximum of 6 funds in such a way that each non-zero weight is at least 0.05. As upper bounds for the weights, we fix 0.25 for the first five (the socially responsible ones) and 0.15 for the conventional ones. These weights allow up to a 75% of conventional funds and up to a 100% of socially responsible funds in each feasible portfolio.

Table 3 Selected funds

#	Name	#	Name
F1	Calvert Large Cap Growth A	F6	BlackRock Index Equity Inv A
F2	Calvert Social Investment Equity A	F7	Dreyfus Appreciation
F3	Domini Social Equity Inv	F8	JPMorgan Equity Index Select
F4	Green Century Equity	F9	Legg Mason Cap Mgmt All Cap B
F5	MMA Praxis Core Stock A	F10	MFS Blended Res. Core Equity A

By observing the efficient frontier, the investor can choose the zone of the plane risk-return he is interested in. Formally, this means to determine the fuzzy set $\tilde{C}$. For this, we fix $(r_0, R_0) = (0.26, 1.98)$ with tolerances $(s_r, s_R) = (0.01, 0.02)$.

To determine an instance of the problem (6), we need to fix the weight w for the social responsibility degree in the goal function. Let us set a quite high value, namely $w = 0.8$ to favor those portfolios being quite far from the efficient frontier if they are good with regard to SRI.

The optimal solution of (6) is the portfolio N1 in Table 4, whose degree of membership of the decision set is 0.6262. With this solution, the investor gets an expected return $r = 0.258$, with a risk $R = 1.98$ and a social responsibility degree $s = 0.3808$. It is interesting to compare this optimal solution with other alternatives, and therefore Table 4 contains the six best portfolios that are optimal with regard to the portfolios with the same composition. Notice that this does not mean that portfolio N2 is the second best solution of (FP), since there are infinitely many portfolios near to N1 that are better than N2. What we can say is that, if we look for a portfolio with a composition different from that of N1, the best possibility is N2, and so on.

Figure 5 shows the position of the portfolios appearing in Table 4 in the risk-return plane. We see that N2 is completely efficient. When compared to N1, it has a similar expected return, a substantially better risk, but a significantly lower social responsibility degree. By contrast, portfolio N3 is again a good solution with regard to social responsibility (it has the second best SRD), but it is worse than N1 because of its SRD, and worse than N2 because of its significantly lower degree of efficiency.

In general, when applying a heuristical procedure for solving a larger instance of (6), it is useful to save not only the best portfolio along the search process, but the best portfolio found for each composition. Hence, in the end we can present the investor not only the optimal portfolio, but a list of alternatives for different compositions. These alternatives are ordered a priori according to his own preferences. In this way the investor is given a last chance to decide which portfolio suits better his preferences with regard to the trade off between risk, return and social responsibility.

Table 4 The six best solutions for different portfolio compositions

	Portfolio										$\mu_{\tilde{D}}$	SRD	Return	Risk
N1	0.23187	0.25	0.218134	0	0	0	0	0	0.1	0.15	0.6262	0.3808	0.268	1.98
N2	0.20585	0.25	0	0.25	0	0	0	0	0.14415	0.15	0.6250	0.2922	0.264	1.931
N3	0.25	0.25	0.15	0.05	0	0	0	0	0.15	0.15	0.6185	0.3476	0.272	1.973
N4	0.16669	0.25	0	0.23331	0	0.05	0	0	0.15	0.15	0.5913	0.2690	0.26	1.924
N5	0.16803	0.25	0	0.23197	0	0	0.05	0	0.15	0.15	0.5818	0.2688	0.26	1.926
N6	0.25	0.25	0.15	0	0	0.05	0	0	0.15	0.15	0.5803	0.3207	0.273	1.978

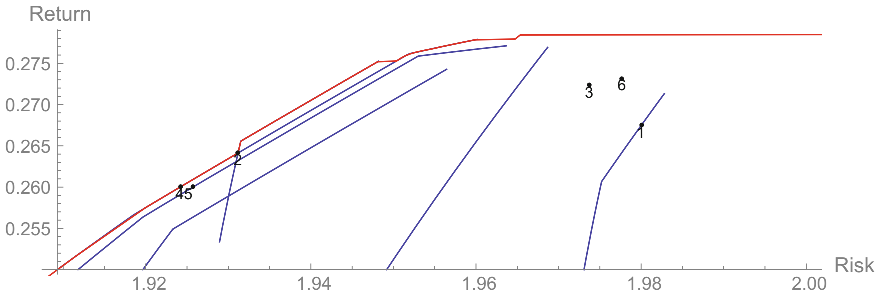


Fig. 5 Location on the plane risk-return of the best solutions

With this proposal, the investor knows exactly what he is missing with respect to the financial goals by accepting the solution of (6), and if he considers the financial cost excessive, he has the possibility of choosing a more conservative alternative among the proposed list or even solving again (FP) with a lower value for the weight of the social responsibility degree.

6 Conclusion

In this chapter we have seen how fuzzy techniques can be applied to the portfolio selection problem in order to deal with different issues related to the subjectivity of the investor's preferences: on one hand, the integrality constraints considered in Sect. 4 make the problem very sensitive to small changes of the risk and return preferences, and our proposed model look for the best solution taking into account that those preferences are soft ones and, hence, the investor will accept slight variations if they provide a reasonable improvement of the solution. On the other hand, when considering non-financial goals as in Sect. 5, our model provides a precise way of prioritizing the financial behavior of the selected portfolio without disregarding its non-financial properties. Of course, it would not be reasonable to expect that a single model would be suitable for reflecting the preferences of every investor (even if it has some adjustable parameters to this end), and hence any other investor's profiles will need essentially different models involving new ideas, and this leaves a rich field for future research.

References

1. Arenas Parra, M., Bilbao Terol, A., Rodríguez Uría, M.V.: A fuzzy goal programming approach to portfolio selection. Eur. J. Oper. Res. **133**, 287–297 (2001)
2. Bundesgesetzblatt: Teil I Nr 62 (2003)

3. Cadenas, J.M., Carrillo, J.V., Garrido, M.C., Ivorra, C., Liern, V.: Exact and heuristic procedures for solving the fuzzy portfolio selection problem. Fuzzy Optim. Decis. Making **11**(1), 29–46 (2012)
4. Calvo, C., Ivorra, C., Liern, V.: Fuzzy portfolio selection with non-financial goals: exploring the efficient frontier. Ann. Oper. Res. (2014). doi:10.1007/s10479-014-1561-2
5. Calvo, C., Ivorra, C., Liern, V.: Fuzzy portfolio selection including cardinality constraints and integer conditions. J. Optim. Theory Appl. **170**(1), 343–355 (2016)
6. Chifu, H., Litzenberg, R.H.: Foundations for Financial Economics. North Holland, Amsterdam (1988)
7. Delgado, M., Verdegay, J.L., Vila, M.A.: Fuzzy linear programming: from classical methods to new applications. In: Delgado, M., Kacprzyk, J., Verdegay, J.L., Vila, M.A. (eds.) Fuzzy Optimization: Recent Advances. Physica Verlag, Heidelberg (1994)
8. Estrada, J.: The cost of equity in emerging markets: a downside risk approach. Emerg. Mark. Q. 19–30 (2000)
9. Estrada, J.: Mean-semivariance behavior: downside risk and capital asset pricing. Int. Rev. Econ. Finance **16**(2), 169–185 (2007)
10. Francis, J.C., Kim, D.: Modern Portfolio Theory: Foundations, Analysis and New Developments. Wiley, New Jersey (2013)
11. Grinold, R., Kahn, R.: Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controling Risk. McGraw Hill, New York (2000)
12. Hallerbach, W., Ning, H., Soppe, A., Spronk, J.: A framework for managing a portfolio of socially responsible investments. Eur. J. Oper. Res. **153**, 517–529 (2004)
13. Horniman, M.D., Jobst, N.J., Lucas, C.A., Mitra, G.: Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints. Quant. Finance **1**, 489–501 (2001)
14. Inuiguchi, M., Tanino, T.: Portfolio selection under independent possibilistic information. Fuzzy Sets Syst. **115**, 83–92 (2000)
15. Jondeau, E., Rockinger, M.: Optimal portfolio allocation under higher moments. Eur. Financ. Manage. **12**, 29–55 (2006)
16. Jorion, P.: Value at Risk: The New Benchmark for Managing Financial Risk, 3rd edn. McGraw-Hill, New York (2006)
17. Joro, T., Na, P.: Portfolio performance evaluation in mean-variance-skewness framework. Eur. J. Oper. Res. **175**(1), 446–461 (2006)
18. Kaplan, P.D., Siegel, B.: Portfolio theory is alive and well. J. Invest. **3**(3), 18–23 (1994)
19. Getting Started With KLD STATS, KLD Research & Analytics, Inc. (2008)
20. Konno, H., Suzuki, K.: A mean-variance-skewness portfolio selection model. J. Oper. Res. Soc. Jpn. **38**, 173–187 (1995)
21. Konno, H., Yamazaki, H.: Mean-absolute deviation portfolio optimization model and its applications to Tokyo Stock Market. Manage. Sci. **37**, 519–531 (1991)
22. Lai, Y.J., Hwang, C.L.: Fuzzy Mathematical Programming: Theory and Applications. Springer, Berlin (1992)
23. León, T., Liern, V., Vercher, E.: Viability of infeasible portfolio selection problems: a fuzzy approach. Eur. J. Oper. Res. **139**, 178–189 (2002)
24. León, T., Liern, V., Vercher, E.: Fuzzy Mathematical Programming for Portfolio Management, Financial Modelling. Physica-Verlag, Heidelberg (2000)
25. León, T., Liern, V., Marco, P., Segura, J.V., Vercher, E.: A downside risk approach for the portfolio selection problem with fuzzy returns. Fuzzy Econ. Rev. **9**(1), 61–77 (2004)
26. Lien, D., Tse, Y.K.: Hedging downside risk: futures vs. options. Int. Rev. Econ. Finance **10**, 159–169 (2001)
27. Markowitz, H.M.: Portfolio selection. J. Finance **7**, 79–91 (1952)
28. Markowitz, H.M.: Portfolio Selection: Efficient Diversification of Investments. Wiley, New York (1959)
29. Ortí, F.J., Sáez, J., Terceño, A.: On the treatment of uncertainty in portfolio selection. Fuzzy Econ. Rev. **7**, 59–80 (2002)

30. Pfleiderer, P.: Is Modern Portfolio Theory Dead? Come On. http://techcrunch.com/2012/08/11/is-modern-portfolio-theory-dead-come-on/ (2012)
31. Roy, A.D.: Safety first and the holding of assets. Econometrica **20**(3), 431–449 (1952)
32. Roy, A.D.: Risk and rank or safety first generalised. Economica **23**(91), 214–228 (1956)
33. Sharpe, W.F.: Capital asset prices: a theory of market equilibrium under conditions of risk. J. Finance **19**(3), 425442 (1964)
34. Sharpe, W.F.: The sharpe ratio. J. Portfolio Manage. **21**(1), 49–58 (1994)
35. Sortino, F.A., Price, L.N.: Performance measurement in a downside risk framework. J. Invest. **3**(3), 59–64 (1994)
36. Stambaugh, F.: Risk and value at risk. Eur. Manage. J. **14**(6), 612–621 (1996)
37. Steuer, R., Qi, Y., Hirschberger, M.: Suitable-portfolio investors, nondominated frontier sensitivity, and the effect of multiple objectives on standard portfolio selection. Ann. Oper. Res. (2007). doi:10.1007/s10479-006-0137-1
38. Taleb, N.N.: The Black Swan: The Impact of the Highly Improbable. Random House (2007)
39. Tanaka, H., Guo, P., Türksen, I.B.: Portfolio selection based on fuzzy probabilities and possibility distributions. Fuzzy Sets Syst. **3**, 387–397 (2000)
40. SIF: 2012 Report on Socially Responsible Investing Trends in the United States. Social Investment Forum, Washington, DC (2012)
41. Watada, J.: Fuzzy portfolio selection and its applications to decision making. Tatra Mount. Math. Publ. **13**, 219–248 (1997)

On Fuzzy Convex Optimization to Portfolio Selection Problem

Ricardo Coelho

Abstract The goal of an investor is to maximize the required return in an investment by minimizing its risk. With this in mind, a set of securities are chosen according to the experience and knowledge of the investor, which subjective evaluations. Selecting these securities is defined as the portfolio selection problem and it can be classified as convex programming problems. These problems are of utmost importance in a variety of relevant practical fields. In addition, since ambiguity and vagueness are natural and ever-present in real-life situations requiring solutions, it makes perfect sense to attempt to address them using fuzzy convex programming technique. This work presents a fuzzy set based method that solves a class of convex programming problems with vagueness costs in the objective functions and/or order relation in the set of constraints. This method transforms a convex programming problem under fuzzy environment into a parametric convex multi-objective programming problem. The obtained efficient solutions to the transformed problem by satisfying an aspiration level defined by a decision maker. This proposed method is applied in a portfolio selection numerical example by using BM&FBOVESPA data of some Brazilian securities.

1 Introduction

Due to desire of maximizing the expected income over a given time horizon, the investors make portfolio and investment decisions. These decisions are supported by the subjective evaluations of income expectations over the chosen time horizon and the risk preferences of the profile of each investor. In this context, we can identify two important problems in achieving the desired objective. The first is the uncertainty in

R. Coelho (✉)
Departamento de Estatística e Matemática Aplicada, Centro de Ciências,
Universidade Federal do Ceará, Av. Mister Hull, s/n, Campus do Pici, Bloco 910,
Fortaleza 60440-900, Brazil
e-mail: rcoelhos@dema.ufc.br

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_8

the subjective evaluation for each scenario because there exist a number of decisions according to the expert's knowledge. The second is the actual decision-making on feasible solutions.

The two problems described in the previous paragraph involve to develop an interface between two close areas, Fuzzy Sets and Systems and Decision Support Systems. On the one hand, in the early sixties, based on the fact that classical logic does not reflect, to the extent that it should, the omnipresent imprecision in the real world, L.A. Zadeh proposed the Theory of Fuzzy Sets and Fuzzy Logic. Nowadays Fuzzy Logic, or rather Soft Computing, is employed with great success in the conception, design, construction and utilization of a wide range of products and systems whose functioning is directly based on the human beings reason ways. On the other hand, the term Decision Support System (DSS) was coined at the beginning of the 70's to feature the computer programs that could support a user in making decisions when facing ill-structured problems. Nowadays, software for supporting decision-making is available for almost any management and optimization problem, that involve minimization (or maximization) of one or various objective functions in a domain that can be constrained or not. If all the functions are linear, we obviously have a linear program. Otherwise, the problem is called a nonlinear program. Nowadays we can use highly efficient and robust algorithms and software for linear programming, which are important tools for solving problems in diverse fields. However, many realistic problems cannot be enough represented or approximated as a linear program owing to the nature of the non-linearity of the objective function and/or the non-linearity of any of the constraints.

As it is well known, convex programming represents a special class of nonlinear programming in which the objective is a convex function or are various convex functions over a convex feasible set. Thus, it is clear that convex programming encompasses all linear problems, including applications in scheduling, planning and flow computations, and they may be used to solve some interesting combinatorial optimization problems. There are several classes of problems that are naturally expressed as convex problems. Examples of such problems can be found in game theory, engineering modeling, design and control, problems involving economies of scale, facility allocation and location problems, problems in microeconomics among others. Several applications and test problems for quadratic programming can be found in [10, 16]. For instance, the risk investment analysis, first introduced by Markowitz [15], is an important research field in the modern finance by modeling this problem as a convex optimization problem.

The paper is organized as follows: Sect. 2 introduces the proposed method to solve convex programming problems under a fuzzy environment. This fuzzy environment can be in the costs of the objective function(s) and/or in the set of constraints. The approach described in this work uses two phases: the first one transforms the problem to be optimized into a parametric convex multi-objective programming problem while in the second part the parametric problem is solved for a satisfaction level given by means of classical optimization techniques. To illustrate the approach Sect. 3 offers a general portfolio selection problem formulated as a fuzzy convex multi-objective programming. Section 4 presents numerical simulations and an analysis of the results obtained. Finally, in Sect. 5, some conclusions are pointed out.

2 Parametric Convex Programming Approach

There are several approaches that solve fuzzy mathematical programming problems, see [1] and [13], which use some defuzzification index, represent the fuzzy coefficients by intervals or transform this fuzzy problem into a parametric mathematical programming problem. The main goal is to transform this imprecise problem into a classical problem and using classical techniques to solve the equivalent problem. In this work, we will focus on the parametric approach in order to transform fuzzy problems into many classical problems with a parameter representing the satisfaction level which belongs to the interval [0,1]. Another way would be defining a new parameter as a new decision variable and to find out the optimal satisfaction level. However, due to space limitations we will not consider this approach.

The parametric approach is divided into two parts: to transform a fuzzy problem into a classical parametric problem; a mathematical formulation of the classical parametric problem which is equivalent to the original fuzzy problem.

2.1 Parametric Approach to Solve Convex Programming Problems Under Fuzzy Environment

The imprecise costs in the objective function(s) of programming problems are represented by fuzzy numbers. These imprecise costs can be defined by the decision maker and they permit some violations that it is not possible in the classical case. Thus, a programming problem with fuzzy costs in the objective function(s) can be formulated in the following way:

$$\begin{array}{ll} \widetilde{\min} & F(\tilde{\mathbf{c}}; \mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in \Omega \end{array} \tag{1}$$

where $F = (f_1, \ldots, f_m)(m \geq 1)$ is an m vector of objective function(s), $\tilde{\mathbf{c}} \in \mathbb{F}(\mathbb{R}^{n\times m})$ represents a vector of fuzzy costs, and $\Omega \subset \mathbb{R}^n$.

A set of membership functions defines the satisfaction level of a feasible solution $\mathbf{x} \in \mathbb{R}^n$. These membership functions can be formulated as follows:

$$\mu_i : \mathbb{R} \to (0, 1], \quad i \in I$$

where μ is a membership function and I is the set that contains all imprecise costs.

A fuzzy number can be represented by a classical one in several ways by using some techniques as the Yager's index [2], the mass center, among many others which are indexes of defuzzification.

In this work, the fuzzy number is transformed into an interval that is defined by lower and upper bounds of the fuzzy number for a determined α-cut level.

Therefore, Problem (1) can be rewritten as:

$$\widetilde{\min}\left\{F([\mathbf{c^L},\mathbf{c^U}]_\alpha;\mathbf{x}) \mid \mathbf{x}\in\Omega\right\}$$

Thus, the fuzzy feasible solution to the original fuzzy problem is a set of optimal solutions for each $\alpha \in (0,1]$. It can be defined in the following way:

$$S(\alpha) = \min\left\{F(\mathbf{c}_\alpha;\mathbf{x}) \mid \mathbf{x}\in\Omega\right\} \tag{2}$$

where $c_\alpha \in [c^L, c^U]_\alpha$ is a real value obtained by a linear combination of bounds of the interval formed by the α-cut level with $\alpha \in (0,1]$.

This fuzzy optimal solution can be found, α-cut by α-cut, by solving the equivalent parametric multi-objective programming problem.

Hence, the fuzzy multi-objective programming problem was parametrized in the end of first part. In the second one the parametric problem is solved for each one of different α values by using conventional multi-objective programming techniques. Thus, the optimal solution achieved by each α of the parametric problem satisfies the Karush-Kuhn-Tucker sufficient optimality conditions to the multi-objective case [4]. This point is described as an efficient solution of original fuzzy problem.

The results for each α value generate a set of satisfaction solutions $S(\alpha)$ and, therefore, according to Representation Theorem to fuzzy numbers, it can be defined as

$$\tilde{S} = \bigcup_\alpha (1-\alpha)S(\alpha),$$

where it can be used to associate all these α-optimal solutions. Hence, it is shown that the feasible solutions are attained by the parametric approach. It determines a fuzzy optimal solution $\tilde{S}$ which solves the original fuzzy multi-objective programming problem.

2.2 *Mathematical Formulation of Parametric Approach to Fuzzy Costs in the Objective Function*

A multi-objective approach that solves a fuzzy linear programming problem with imprecise costs in the objective functions is presented in [8, 9]. This approach can be extended to solve nonlinear programming problems with one or several objectives. In [11], another multi-objective approach is developed that solves nonlinear programming problems with only one objective with imprecise coefficients in the objective function and in the set of constraints. But it can be extended to solve fuzzy multi-objective problems too.

In this work, multi-objective programming problems with fuzzy costs in the objective functions are formulated in the following way:

$$\begin{array}{ll} \min & \left[f_1(\tilde{\mathbf{c}}_1;\mathbf{x}),\ldots,f_m(\tilde{\mathbf{c}}_m;\mathbf{x})\right] \\ \text{s.t.} & \mathbf{x}\in\Omega \end{array} \tag{3}$$

where $\mathbf{x}$ is an n vector of real numbers, $\tilde{\mathbf{c}}_i$ is a vector of fuzzy numbers with p_i components, $i=\{1,\ldots,m\}$. The fuzzy numbers are characterized by the membership functions that are defined by the decision maker. The membership functions are defined as

$$\mu_j : \mathbb{R}\rightarrow[0,1],\ \ j\in\mathbb{J}=\{1,2,\ldots,m\}$$

In particular, these membership functions are described by:

$$\mu_j(y)=\begin{cases} 0 & \text{if } c_j^U < y \text{ or } y < c_j^L \\ L_j(y) & \text{if } c_j^L \le y \le c_j^1 \\ R_j(y) & \text{if } c_j^2 \le y \le c_j^U \end{cases} \qquad j\in\mathbb{J} \tag{4}$$

where $\mathbf{L}(\cdot)$ and $\mathbf{R}(\cdot)$ are strictly increasing and decreasing continuous functions, respectively, $L_j(c_j^1)=R_j(c_j^2)=1,\ j\in\mathbb{J}$.

The problem considered in Verdegay [8] presents trapezoidal membership functions to describe the fuzzy numbers. In this work, we consider it as defined in (4).

By considering the $(1-\alpha)$-cut level of all the costs, for $\alpha\in[0,1]$, $\forall\, x\in\mathbb{R}$

$$\mu_j(x)\ge 1-\alpha \Leftrightarrow L_j^{-1}(1-\alpha)\le x\le R_j^{-1}(1-\alpha),\ \ j\in\mathbb{J}$$

Thus, according to the parametric transformations shown above, a fuzzy solution to (1) can be obtained from a parametric solution of a equivalent parametric multi-objective programming problem which is formulated as:

$$\begin{array}{ll} \min & \left[f_1(\mathbf{c_1^1};\mathbf{x}), f_1(\mathbf{c_1^2};\mathbf{x}),\ldots,f_1(\mathbf{c_1^{2^{p_1}}};\mathbf{x}),\ldots,\right. \\ & \left. f_m(\mathbf{c_m^1};\mathbf{x}), f_m(\mathbf{c_m^2};\mathbf{x}),\ldots,f_m(\mathbf{c_m^{2^{p_m}}};\mathbf{x})\right] \\ \text{s.t.} & \mathbf{x}\in\Omega,\ \ \mathbf{c^k}\in\mathbb{E}(1-\alpha), \\ & \alpha\in[0,1],\ \ k=1,2,\ldots,2^{p_j}, \end{array} \tag{5}$$

where $\mathbb{E}(1-\alpha)$, for each $\alpha\in[0,1]$, is the set of vectors in $\mathbb{R}^{p_j}$, such that p_j informs the amount of fuzzy numbers that represent the imprecise costs in each objective function, for $j\in\{1,2,\ldots,m\}$.

Each $(1-\alpha)$-cut level element of this set is in the lower bound, $L_j^{-1}(1-\alpha)$, or in the upper bound, $R_j^{-1}(1-\alpha)$, i.e., $\forall\, k = 1, 2, \ldots, 2^{p_j}$,

$$\mathbf{c^k} = (c_1^k, c_2^k, \ldots, c_m^k) \in \mathbb{E}(1-\alpha)$$

where c_j^k is equal to $L_j^{-1}(1-\alpha)$ or $R_j^{-1}(1-\alpha)$, for all $j = 1, \ldots, m$

It is clear that a parametric optimal solution to (5) is part of the fuzzy optimal solution to (3). This approach was developed to the convex case with one only objective, as described in [5, 18, 20], and here is extended to the convex case with several objectives. The parametric optimal solutions can be obtained by using any optimization method to solve classical multi-objective programming problems.

2.3 Mathematical Formulation of Parametric Approach to Fuzzy Order Relation in the Set of Constraints

As in [6, 17, 19], the constraints of a convex problem are defined as having a fuzzy nature, that is, some violations in the accomplishment of such restrictions are permitted. Thus, the convex problem can be addressed as follows

$$\begin{aligned} &\min\ F([\mathbf{c^L}, \mathbf{c^U}]_\alpha; \mathbf{x}) \\ &\text{s.t.}\ \ g_i(\mathbf{x}) \leq^f b_i, i = 1, \ldots, m \\ &\qquad x_j \geq 0, j = 1, \ldots, n \end{aligned} \tag{6}$$

where g_i, for each $i = 1, \ldots, m$, is a convex function by building a convex constraint set. Besides, the membership functions

$$\mu_i : \mathbb{R}^n \to [0, 1], \quad i = 1, \ldots, m$$

on the fuzzy constraints are to be determined by the decision maker. It is clear that each membership function will give the membership (satisfaction) degree such that any $x \in \mathbb{R}^n$ accomplishes the corresponding fuzzy constraint upon which it is defined. These membership functions can be formulated as follows

$$\mu_i(\mathbf{x}(\lambda)) = \begin{cases} 1 & g_i(\mathbf{x}) \leq b_i \\ 1 - \dfrac{g_i(\mathbf{x}) - b_i}{d_i} & b_i \leq g_i(\mathbf{x}) \leq b_i + d_i \\ 0 & g_i(\mathbf{x}) > b_i + d_i \end{cases}$$

with $i = 1, \ldots, m$. In order to solve this problem in a two-phase method, first let us define for each fuzzy constraint,

$$X_i = \left\{\mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \leq^f b_i, \mathbf{x} \geq 0\right\}.$$

If $\mathbf{X} = \bigcap X_i (i = 1, \dots, m)$ then the former fuzzy quadratic problem can be addressed in a compact form as

$$\min \left\{ F([\mathbf{c}^{\mathbf{L}}, \mathbf{c}^{\mathbf{U}}]_\alpha; \mathbf{x}) \mid \mathbf{x} \in \mathbf{X} \right\}.$$

It is clear that $\forall \lambda \in (0, 1]$, an λ-cut of the fuzzy constraint set will be the classical set

$$X(\lambda) = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \mu_X(\mathbf{x}) \geq \lambda \right\}$$

where $\forall \mathbf{x} \in \mathbb{R}^n$,

$$\mu_X(\mathbf{x}) = \min \mu_i(\mathbf{x}(\lambda)), i = 1, \dots, m$$

Hence an λ-cut of the i-th constraint will be denoted by $X_i(\lambda)$. Therefore, if $\forall \lambda \in (0, 1]$,

$$S(\lambda) = \left\{ \mathbf{x} \in \mathbb{R}^n \mid F([\mathbf{c}^{\mathbf{L}}, \mathbf{c}^{\mathbf{U}}]_\alpha; \mathbf{x}) = \min \ F([\mathbf{c}^{\mathbf{L}}, \mathbf{c}^{\mathbf{U}}]_\alpha; \mathbf{y}), \ \mathbf{y} \in X(\lambda) \right\}$$

the fuzzy solution to the problem will therefore be the fuzzy set defined by the following membership function

$$S(\mathbf{x}) = \begin{cases} \sup\{\lambda : \mathbf{x} \in S(\lambda)\} & \mathbf{x} \in \bigcup_\lambda S(\lambda) \\ 0 & \text{otherwise.} \end{cases}$$

Provided that $\forall \lambda \in (0, 1]$,

$$X(\lambda) = \bigcap_{i=1,\dots,m} \{\mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \leq r_i(\lambda), \mathbf{x} \geq 0, \mathbf{x} \in \mathbb{R}^n\}$$

with $r_i(\lambda) = b_i + d_i(1 - \lambda)$. The operative solution to the former problem can be found, λ-cut by λ-cut, by means of the following auxiliary parametric convex programming model,

$$\begin{array}{ll} \min & F([\mathbf{c}^{\mathbf{L}}, \mathbf{c}^{\mathbf{U}}]_\alpha; \mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \leq b_i + d_i(1 - \lambda), i = 1, \dots, m \\ & x_j \geq 0, j = 1, \dots, n, \lambda \in [0, 1]. \end{array} \tag{7}$$

Thus, the fuzzy convex programming problem was parameterized at the end of the first phase.

In the second phase the parametric quadratic programming problem is solved for each of the different λ values using the technique described in previous subsection.

3 Portfolio Selection Problem

In order to illustrate the above described parametric method to solve fuzzy convex programming problems, we are going to focus on general portfolio problems. It is important to emphasize that, at the present time, we do not try to improve other solution methods for this kind of important problems, but only to show how our solution approach performs. In [15], a description of a classical portfolio selection problem is given that was formulated by Markowitz as a convex programming problem.

Markowitz's model combines probability and optimization techniques to model the behavior of investment under uncertainty. The investors are assumed to strike a balance between minimizing the risk and maximizing the return of their investment. The risk is characterized by the variance, and the return is quantified by the mean, of a portfolio of assets. The two objectives of an investor are thus to minimize the variance of a portfolio and to maximize the expected value of return.

Markowitz model for portfolio selection can be formulated mathematically in two ways: minimizing risk when a level of return is given and maximizing return when a level of risk is given. Hence, assume that there are n securities denoted by S_j ($j = 1, \dots, n$), the former problem is formulated on the following way:

$$\begin{aligned} \min\ & \mathbf{x^t Q x} \\ \text{s.t. } & \mathbf{x^t E(R)} \geq \rho \\ & \mathbf{1x} = 1 \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{8}$$

where $\mathbf{x}$ is an n vector that represents the percentage of money invested in asset, i.e., the proportion of total investment funds devoted to each security; $\mathbf{E(R)}$ is the average vector of returns over m periods because $\mathbf{R} = [r_{ij}]$ is an $m \times n$ matrix that represents the random variables of the returns of asset varying in m discrete times; ρ is a parameter representing the minimal rate of return required by an investor; and $\mathbf{Q} = [\sigma_{ij}^2]$ is a covariance $n \times n$ matrix between returns of asset which can be written as

$$\sigma_{ij}^2 = \sum_{k=1}^{m} \frac{\left(r_{ki} - E(r_i)\right)\left(r_{kj} - E(r_j)\right)}{m-1}. \tag{9}$$

On the other hand, the latter problem is formulated as

$$\begin{aligned} \min\ & \mathbf{x^t E(R)} \\ \text{s.t. } & \mathbf{x^t Q x} \leq \gamma \\ & \mathbf{1x} = 1 \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{10}$$

where γ is the maximum risk level the investor would bear.

The expected return rate, ρ, and the maximum risk level, γ, are decision maker's values that represent an expert's knowledge. These two formulations of portfolio selection problems can be mixed and formulated as a bi-objective convex programming problem. Moreover, uncertainty and multiple objectives are the important factors in decision making. From a practical viewpoint, it is usually difficult to determine exactly the coefficients in mathematical programming problems due to various kinds of uncertainties. However, it is sometimes possible to estimate the perturbations of coefficients by intervals, fuzzy numbers or possibilistic distributions.

The portfolio problem is a typical decision making problem under uncertainty which has received considerably attention in the literature recently. This problem addresses the dilemma that each investor faces the conflicting objectives of high profit versus low risk. In this work, the uncertainties in the objective functions are represented by fuzzy costs and the order relation in the set of constraints is fuzzy too. So, portfolio selection problem can be formulated in the following way

$$\begin{aligned} \min\ & \mathbf{x^t E(\tilde{R})} \\ \text{s.t. } & \mathbf{x^t Q x} \leq^f \gamma \\ & \mathbf{1x} = 1 \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{11}$$

There are many formulations to describe a portfolio selection problem that are more realistic than one was presented in this work. One of them is to put another objective function called "Value at Risk", as described in [12]. It is defined as a threshold value, which is a given probability level of the worse loss on the portfolio over the given time horizon. However, we choose this convex formulation to show the efficiency of our approach, and we will extend it to apply in other formulations in the next step.

4 Numerical Experiments

In this section, a portfolio selection problem with fuzzy costs and fuzzy order relation in the set of constraints are analyzed. In Sect. 4.1 we will show the used data to formulate the fuzzy portfolio problems. Then in Sect. 4.2 the computational results and a comparative analysis of the classic and parametric approaches responses will be presented.

The tests were all performed on a PC with 2.7 GHZ Intel® Core™ i7, 16 GB RAM running MacOS Sierra operational system. All the problems presented in this work were resolved using **fmincon** function to solve constraint programming problems of *ToolBox Optimization* of MATLAB® R2015a program. The evolutionary algorithm parameter are 100 generations and 100 individuals in the population, while the crossover and mutation index are 0.6 and 0.1, respectively.

4.1 Formulation of the Numerical Examples

In order to show the performance of our method, we used the set of historical data shown in Table 1 took by BM&FBOVESPA which is an important market for Brazilian securities. It was chosen ten Brazilian securities and the columns 2–11 represent Cemig, Cesp, Copel, Eletrobras, Embraer, Light, Petrobras, Unipar, Usiminas, and Vale securities data, respectively. The returns on the ten securities, during the period of 1994 up to 2016, are presented in Table 1.

The vagueness was inserted into the set of constraints in the form around 40% variation in the modal value of each constraint function. Besides, each component of the vector of imprecise costs is a fuzzy number and they are transformed into a interval.

This example will consider performance of portfolios with respect to “return” and “risk”. This assumes that a euro of realized or unrealized capital gains is exactly equivalent to an euro of dividends, no better and no worse. This assumption is appropriate for certain investors, for example, some types of tax-free institutions. Other ways of handling capital gains and dividends, which are appropriate for other investors, can be viewed in [15].

4.2 Results and Analysis

Here we show two results obtained for the portfolio selection problem with uncertainties by the fuzzy convex programming method introduced in Sect. 2. The problems described in this work were solved by using the equivalent parametric multi-objective problem as presented by Problems (5) and (7). The data from BM&FBOVESPA shown in Table 1 was used on two ways of the Problem (11): (i) only order relation in the set of constraints is uncertain; and (ii) the linear objective function which maximizes the return has imprecise costs and the order relation in the set of constraints is uncertain.

Table 2 presents the solutions of the parametric portfolio selection problem with ten securities and this same problem with imprecise order relation in the set of constraints. It is clear to see that the investor should choose only four securities of them for any satisfaction level. With this choice, the expected value of return of this investment is between 22% up to 26%. To highlight, the maximum risk allowance is 18% and the maximum tolerance is 7%.

Figure 1 presents the solutions of the parametric portfolio selection problem with ten securities and this same problem with imprecise order relation in the set of constraints. The line represents the interpolation of the obtained points for each satisfaction level.

Figure 2 presents the solutions of the parametric multi-objective portfolio selection problem with ten securities and this same problem with imprecise costs in the linear objective function. Besides, the order relation in the set of constraints is impre-

Table 1 BM&FBOVESP data from 1995 to 2016

Periods	#1 Cemig	#2 Cesp	#3 Copel	#4 Eletrobras	#5 Embraer	#6 Light	#7 Petrobras	#8 Unipar	#9 Uniminas	#10 Vale
1994	0.7651	0.0000	0.5017	0.4375	0.6667	0.3600	0.1207	0.4875	0.0993	1.8763
1995	−0.7856	−0.9800	−0.2342	−0.1204	−0.9050	0.0163	−0.3846	−0.8386	−0.4839	−0.1147
1996	1.0674	1.0545	0.5942	0.4144	1.1053	0.1865	2.1250	−0.3582	0.3125	-0.8907
1997	−0.0226	0.1945	0.1364	-0.8508	1.3505	0.2601	0.4800	−0.3721	8.0476	−0.2185
1998	−0.6282	−0.6499	−0.4480	−0.6252	−0.7788	−0.6839	−0.6000	0.1111	−0.5789	−0.3697
1999	0.6903	−0.5608	0.6826	0.8995	−0.1346	0.3537	4.4050	3.3333	0.5000	2.1579
2000	−0.0420	−0.1084	0.1189	−0.0876	0.3722	0.1256	−0.8787	−0.1308	−0.0500	0.1071
2001	0.3416	0.7838	0.2163	−0.0763	0.0130	−0.4531	0.0823	−0.0265	−0.0877	0.118
2002	−0.2540	−0.5114	−0.4937	−0.2913	−0.0120	−0.6177	0.0059	0.2455	0.2212	0.9615
2003	0.4489	0.7674	0.0813	0.8538	0.5372	0.7784	0.5928	0.5328	2.9370	0.5906
2004	0.6705	−0.1053	0.1329	−0.1200	−0.1684	−0.2480	0.2663	1.0905	0.9996	−0.5346
2005	0.4065	0.0245	0.5255	−0.0130	0.1392	−0.7549	−0.6122	−0.4579	0.0262	0.2649
2006	0.1400	0.8660	0.4381	0.3445	0.2250	0.0853	0.3194	−0.0966	0.8713	−0.3330
2007	−0.6295	0.8103	0.3721	−0.5316	−0.0862	0.7197	0.9270	0.3814	−0.1354	−0.0689
2008	−0.2587	−0.6941	−0.2542	0.0819	−0.5628	−0.2370	−0.7382	−0.6633	−0.6886	−0.5331
2009	−0.0259	0.7963	0.6591	0.4036	0.0795	0.1885	0.5151	1.2000	0.9377	0.7876
2010	−0.1496	0.1289	0.0411	−0.3880	0.2408	−0.0212	−0.2665	−0.6818	−0.5738	0.1178
2011	0.3157	0.2466	−0.1311	−0.1978	−0.0034	0.1325	−0.2471	−0.3857	−0.1967	−0.2870
2012	−0.1978	−0.3960	−0.2368	−0.6452	0.2287	−0.2250	−0.1500	0.0698	−0.2029	0.0717
2013	−0.3516	0.2068	−0.1151	−0.0727	0.3073	−0.0090	−0.1821	0.3043	−0.0929	−0.1554
2014	0.5775	0.1256	0.1166	−0.0119	0.2938	−0.2306	−0.4003	−0.1167	−0.0081	−0.3864
2015	−0.7125	−0.5446	−0.3574	−0.0069	0.2353	−0.4183	−0.1064	0.1132	−0.6732	−0.4053
2016	0.2189	0.2255	0.1925	2.9601	−0.4700	0.7535	0.9767	0.2733	1.0547	0.9708

Table 2 Solutions of the parametric portfolio selection problem

λ	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	FunObj
0.0	0.0000	0.0000	0.5893	0.0000	0.0000	0.0000	0.0000	0.1475	0.1636	0.0996	0.2581
0.1	0.0000	0.0000	0.5981	0.0000	0.0000	0.0000	0.0000	0.1446	0.1594	0.0979	0.2548
0.2	0.0000	0.0000	0.6070	0.0000	0.0000	0.0000	0.0000	0.1416	0.1552	0.0962	0.2514
0.3	0.0000	0.0000	0.6163	0.0000	0.0000	0.0000	0.0000	0.1385	0.1508	0.0944	0.2479
0.4	0.0000	0.0000	0.6258	0.0000	0.0000	0.0000	0.0000	0.1353	0.1463	0.0926	0.2444
0.5	0.0000	0.0000	0.6356	0.0000	0.0000	0.0000	0.0000	0.1321	0.1416	0.0907	0.2407
0.6	0.0000	0.0000	0.6457	0.0000	0.0000	0.0000	0.0000	0.1287	0.1368	0.0888	0.2368
0.7	0.0000	0.0000	0.6531	0.0000	0.0000	0.0000	0.0000	0.1254	0.1321	0.0868	0.2328
0.8	0.0000	0.0000	0.6421	0.0000	0.0000	0.0000	0.0000	0.1231	0.1298	0.0853	0.2287
0.9	0.0000	0.0000	0.6303	0.0000	0.0000	0.0000	0.0000	0.1209	0.1274	0.0837	0.2246
1.0	0.0000	0.0000	0.6185	0.0000	0.0000	0.0000	0.0000	0.1186	0.1250	0.0821	0.2203

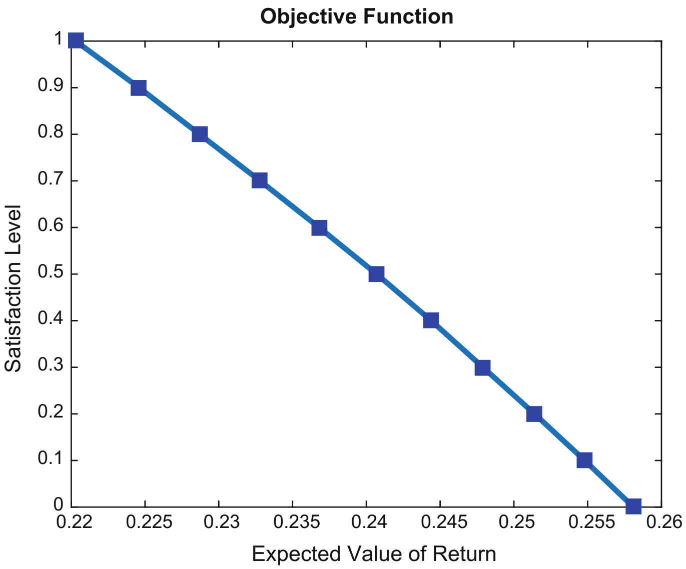


Fig. 1 The fuzzy solution applied in the linear objective function

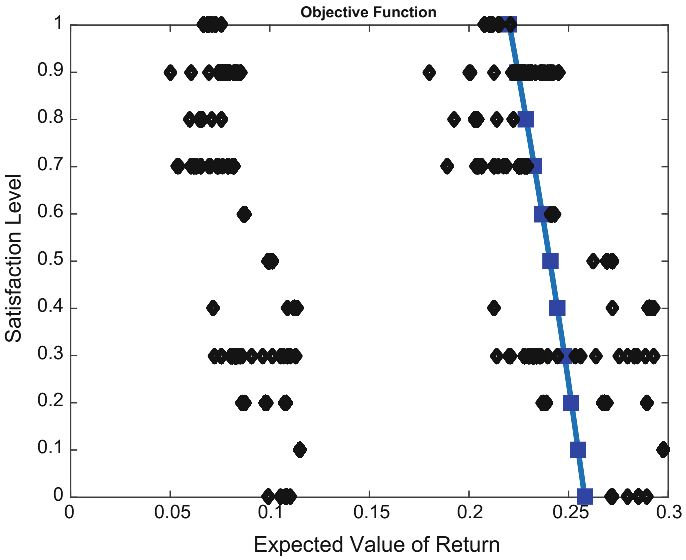


Fig. 2 The fuzzy solution applied in the linear objective function with fuzzy costs

cise too. In this case it is clear to identify the belt that represents the fuzzy efficient solutions and the classical efficient solutions are inside of this belt.

5 Conclusion

Convex Programming problems are very important in a variety of both theoretical and practical areas. When real-world applications are considered, the vagueness appears in a natural way, and hence it makes perfect sense to think in Fuzzy Convex Programming problems. In contrast to what happens with Fuzzy Linear Programming problems, unfortunately until now no solution method has been found for this important class of problems. In this context this paper has presented an operative and novel method for solving Fuzzy Convex Multi-Objective Programming problems which is carried out by performing two phases which finally provide the user with a fuzzy solution. The method has been validated by solving a portfolio selection problem. The obtained solutions allow the author to follow along this research line trying to solve real problems in practice, in such a way that oriented Decision Support Systems involving Fuzzy Convex Programming problems can be built.

An evolutionary algorithm called NSGA-II was used and it produces a sequence of points according to a prescribed set of instructions, together with a termination criterion. Usually we look for a sequence that converges to a set of efficient solutions, but in many cases however we have to be satisfied with less favorable solutions. Then the procedure may stop either (1) if a point belonging to a prefixed set (the solution set) is reached, or (2) if some prefixed condition for satisfaction is verified. In any case, assuming that a solution set is prefixed, the algorithm would stop if a point in that solution set is reached. Frequently, however, the convergence to a point in the solution set is not easy because, for example, of the existence of local optimum points. Hence we must redefine some rules to finish the iterative procedure.

Hence the control rules of the algorithms solving convex programming problems could be associated to the solution set, and to the criteria for terminating the algorithm. As it is clear, fuzziness could be introduced in both points, not assuming it as inherent in the problem, but as help for obtaining, in a more effective way, some solution for satisfying the decision-maker's wishes. This mean that the decision maker might be more comfortable obtaining a solution expressed in terms of satisfaction instead of optimization, as it is the case when fuzzy control rules are applied to the processes.

Acknowledgements The authors want to thank the support provided by the Brazilian agency CNPq with process number 4849002/2013-0.

References

1. Bector, C.R., Chandra, S.: Fuzzy Mathematical Programming and Fuzzy Matrix Games. Studies in Fuzziness and Soft Computing, vol. 169. Springer, Berlin (2005)
2. Bortolan, G., Degani, R.: A review of some methods for ranking fuzzy subsets. Fuzzy Sets Syst. **15**, 1–19 (1985)
3. Carlsson, C., Fullér, R., Majlender, P.: A possibilistic approach to selecting portfolio with highest utility score. Fuzzy Sets Syst. **131**, 13–21 (2002)
4. Chankong, V., Haimes, Y.Y.: Multiobjective Decision Making: Theory and Methodology. North Hollando Series in System Science and Engineering, vol. 8. North Holland, New York, USA (1983)
5. Cruz, C., Silva, R.C., Verdegay, J.L.: Fuzzy costs in quadratic programming problems. Fuzzy Optim. Decis. Making **12**, 231–248 (2013)
6. Cruz, C., Silva, R.C., Verdegay, J.L.: Extending and relating different approaches for solving fuzzy quadratic problems. Fuzzy Optim. Decis. Making **10**, 193–210 (2011)
7. Deb, K.: Multi-objective Optimization Using Evolutionary Algorithms. Wiley, Chichester, UK (2001)
8. Delgado, M., Verdegay, J.L., Vila, M.: Imprecise costs in mathematical programming problems. Control Cybern. **16**(2), 113–121 (1987)
9. Delgado, M., Verdegay, J.L., Vila, M.: Relating different approaches to solve linear programming problems with imprecise costs. Fuzzy Sets Syst. **37**, 33–42 (1990)
10. Floudas, C.A., Pardalos, P.M., Adjiman, C., Esposito, W.R., Gümüs, Z.H., Harding, S.T., Klepeis, J.L., Meyer, C.A., Schweiger, C.A.: Handbook of Test Problems in Local and Global Optimization. Nonconvex Optimization and Its Applications, vol. 33. Kluwer Academic Publishers, Dordrecht (1999)
11. Jiménez, F., Cadenas, J., Sánchez, G., Gómez-Skarmeta, A., Verdegay, J.L.: Multi-objective evolutionary computation and fuzzy optimization. Int. J. Approx. Reason. **43**(1), 59–75 (2006)
12. Jorion, P.: Value at Risk: The New Benchmark for Managing Financial Risk, 3rd edn. McGraw-Hill (2006)
13. Lai, Y.J., Hwang, C.L.: Fuzzy Mathematical Programming: Methods and Applications. Lecture Notes in Economics and Mathematical Systems, vol. 394. Springer, Berlin (1992)
14. León, T., Liern, V., Vercher, E.: Viability of infeasible portfolio selection problems: a fuzzy approach. Eur. J. Oper. Res. **139**, 178–189 (2002)
15. Markowitz, H.M.: Portfolio Selection: Efficient Diversification of Investments, 2nd edn. Blackwell Publisher, Massachusetts, USA (1991)
16. Schittkowski, K.: More Test Examples for Nonlinear Programming Codes. Spring (1987)
17. Silva, R.C., Cruz, C., Verdegay, J.L., Yamakami, A.: A survey of fuzzy convex programming models. In: Lodwick, W.A., Kacprzyk, J. (eds.) Fuzzy Optimization: Recent Advances and Applications. Studies in Fuzziness and Soft Computing, vol. 254, pp. 127–143. Springer, Berlin (2010)
18. Silva, R.C., Cruz, C., Yamakami, A.: A parametric method to solve quadratic programming problems with fuzzy costs. In: IFSA/EUSFLAT 2009. Lisbon, Portugal, July 2009
19. Silva, R.C., Verdegay, J.L., Yamakami, A.: Two-phase method to solve fuzzy quadratic programming problems. In: 2007 IEEE International Fuzzy Systems Conference, London, UK, July 2007
20. Silva, R.C., Verdegay, J.L. Yamakami, A.: A parametric convex programming approach applied in portfolio pelection problem with fuzzy costs. In: 2010 IEEE International Fuzzy Systems Conference, Barcelona, Spain, July 2010
21. Tanaka, H., Guo, P., Türksen, B.: Portfolio selection based on fuzzy probabilities and possibity distributions. Fuzzy Sets Syst. **111**, 387–397 (2000)

Digital Coaching for Real Options Support

Christer Carlsson

Abstract Classical management science is making the transition to analytics, which has the same agenda to support managerial planning, problem solving and decision making in industrial and business contexts but is combining the classical models and algorithms with modern, advanced technology for handling data, information and knowledge. In work with managers in the forest industry, we found out that there is a growing interest to replace the classical net present value (NPV) with real options theory, especially for strategic issues and uncertain, dynamic environments. Uncertainty and dynamics motivate the use of soft computing, i.e. versions of the real options methods that use fuzzy numbers (intervals), macro heuristics, approximate reasoning and evolutionary algorithms. In general, managers can follow the logic of the real options theory but the methods require rather advanced levels of analytics; when the methods are implemented, they will be used by growing numbers of people with more of a business than analytics background. They find themselves in trouble pretty quickly as they need to master methods, they do not fully understand and details of which they forget from time to time. We propose that digital coaching is a way to guide and support users to give them better chances for effective and productive use of real options methods.

Keywords Digital coaching · Analytics · Fuzzy real options

1 Introduction

This chapter has a history and a reason that bridges the past, the present and the future. The history is a paper I wrote called *On the Relevance of Fuzzy Sets in Management Science Methodology* in 1984 [4]. This was a time when we tried to make the case for fuzzy sets in management science and as a support theory for managers who plan the future, and solve problems and make decisions in their daily

C. Carlsson (✉)
IAMSR/Abo Akademi University, Turku, Finland
e-mail: christer.carlsson@abo.fi

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_9

activities. This is an activity where Professor Jose Luis "Curro" Verdegay has made significant contributions over many years and where he is a much sought-after collaborator for many researchers.

If we continue the history a bit, a first version of the paper had been presented and discussed at the 11th meeting of the EURO Working Group on Fuzzy Sets (in which Curro was an active collaborator) at the European Institute for Advanced Studies in Management in Brussels on February 19–20, 1981. The EIASM is the centre for serious research on management in Europe and getting an invitation to run a workshop on fuzzy sets took some negotiation; I was chairing the EURO WG in that period and had to do the negotiating.

Management science methodology—and especially operations research that applied the same methodology for engineering problems and theory development—had already in 1981 been under attack for more than a decade for failing to deal with the real world problems managers have to tackle, for oversimplifying problems and for spending too much time with mathematically interesting but practically irrelevant problems and solutions. The message was simply that management science methodology produced theory and methods that were irrelevant for handling actual management problems. The paper in 1984 argued that fuzzy sets when properly worked into management science methodology would make the models, the algorithms and the theory more relevant and better suited to deal with management problems in practice.

Now, more than 30 years later, we have to admit that we were not successful in bringing it about, that fuzzy sets remained a marginal development in management science and that we have been able to get fuzzy sets based methods accepted only for some limited applications, such as multiple criteria optimisation, real options valuation, logistics optimisation, etc. for which there have been algorithmic benefits of allowing the use of fuzzy numbers.

Management science and operations research have also changed over the decades; two major organisations in the field—TIMS and ORSA—merged and became INFORMS to combine the applications oriented research (TIMS) with the algorithms and theory oriented research (ORSA); now the annual INFORMS conferences collect 2–3000 participants; in Europe the EURO Association is a sister organisation to INFORMS and the annual EURO conferences also collect 2–3000 participants. Both organisations run major, well-established journals with high impact factors and there are dozens of journals publishing material produced under guidance of management science methodology. The field is alive and well and promotes lively research that activates thousands of researchers. The context is there; what is needed for fuzzy sets to be relevant (again) for the research that is carried out—a question and a goal we have had in the International Fuzzy Systems Association (IFSA) for the last decade? Or to be more focused and specific—is there any management science problem area where fuzzy sets theory could be vital for breakthrough research (cf. [15])?

1.1 Analytics

Operations Research and Management Science are now in the process of being transformed by (Business) *Analytics* which is getting the attention of major corporations and senior management. On our part, in our work with complex, difficult problems for large industrial corporations, we have for a number of years been promoting *Soft Computing* to the same audience instead of trying to explain fuzzy sets theory and fuzzy logic in the way it was originally done. The experience we have—summarized in a few words—is that *analytics* methods which implement *soft computing* theory and algorithms are turning out to be very effective and useful for planning, problem solving and decision making in "big data" environments; the "big data" is one of the challenges of the modern digital economy and for which we propose that fuzzy sets would offer instruments for breakthrough research [15].

Analytics adds value to management; it promotes data-driven and analytical decision making, which was somehow "reinvented" as being important and useful for management that had relied on other schools of thought for a couple of decades. Analytics builds on recent software improvements in information systems that has made data, information and knowledge available in real time in ways that were not possible for managers only a few years ago [18]. INFORMS gradually found out that the new movement represents both "potential opportunities" and "challenges" to management science and operations research professionals [25]. The methods and the application cases worked out in the Davenport-Harris book [18] are very close to traditional management science methodology, actually so close that a manager probably fails to see any differences, which is why INFORMS finds "challenges".

Soft Computing (introduced by Lotfi Zadeh in 1991) builds on fuzzy sets theory [31], fuzzy logic, optimisation, neural nets, evolutionary algorithms, macro heuristics and approximate reasoning. Soft Computing is focused on the design of intelligent systems to process uncertain, imprecise and incomplete information; soft computing methods applied to real-world problems offer more robust, tractable and less costly solutions than those obtained by more conventional mathematical techniques.

Liberatore and Luo [25] list four factors that drive the analytics movement: (i) availability of data (ii) improved analytical software (iii) the adoption of a process orientation by organisations, and (iv) managers and executives who are skilled users of information and communication technology. Compared to the experience of the 1980s the last factor is probably the most important driver—there is a new generation of managers and executives in charge of the corporations that are using information technology as part of their daily routines. They work with data, information and knowledge on a real time basis and they continuously hunt for improved analytical tools to help give them competitive advantages. They do not necessarily recognize the analytical tools as classical management science algorithms; analytical software (cf. (ii)) has become user-friendly with graphical user interfaces and visualisation of results; users typically do not have the mathematical background to get into details with the algorithms. Information technology has made data available on a real time basis which allows online planning, problem solving and decision making. Maybe "allow" is not the right verb as online management work in real time is more of a necessity to keep up with the competition. The same driver also explains the adoption of a process orientation (cf. (iii)) as management work typically is group—and teamwork online and in real time. Davenport and Harris [18] describe analytics as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models and fact-based management to drive decisions and actions". Liberatore and Luo [25] identify three levels of modelling—descriptive, predictive and prescriptive—and state that management science and operations research typically would focus on advanced analytics, i.e. prescriptive modelling. They also point out that analytics would focus on the transformation of data into actions through analysis and insight, which in their discussion contributes to the application cases of management science.

The modern movement of *analytics* appears to offer interesting possibilities and opportunities for soft computing; the movement is data-driven which will require tools for handling *imprecision*; the movement is focused on managers who need to deal with real world problems, for which available data, information and knowledge are incomplete, imprecise and uncertain and should allow for fast, often intuitive conclusions; the movement builds on improved analytical software that offers platforms for a multitude of algorithms, intelligent technologies, soft computing, computational intelligence, etc. Modern analytics offers platforms and environments for *digital coaching* of managers in planning, problem solving and decision making.

There are benefits of having worked with management science for a few decades (like myself and my contemporaries)—there has been hundreds of innovative ideas and some successful solutions from which it has been possible to extract working principles and growing understanding of how good science can guide and contribute to successful planning, problem solving and decision making. In the context of the digital economy these processes—not surprisingly—offer new challenges: real-time management challenged by "big data" and relying on fast processing by advanced analytics methods would best be carried out by postdoc-qualified managers—these are rather scarce and would most often be filtered out by corporate career qualifying processes much before they reach senior management positions. Thus there will be a need to reinstate "coaching" functions with the advanced analytics methods to tell/explain to the users what can/should be done, how it should be carried out, what the results are and what they mean, and how they should be applied (with explanations of what could/should not be done).

My storyline is worked out in the context of analytics and soft computing and the history that has formed that context over the last few decades. I will work with fuzzy real options modelling, that is one of the more advanced analytics methods with a theory that is not easily introduced to managerial decision makers but which shows results that represent insight that can offer sustainable competitive advantages. The fuzzy real options modelling is introduced in Sect. 2 and the principles and the state of the art of digital coaching in Sect. 3; then we will use the methods in Sect. 4 to find out how users of real options models could be given coaching support for strategic planning and effective decision making. The chapter will finish with a summary of the main points and with some linking to insight developed by Kahneman [24].

2 Fuzzy Real Options Modelling

Black and Scholes (cf. [17]) introduced the options theory to decide the price for options on financial assets that would be effective in the financial market; Merton (cf. [17]) then proved that similar options modelling also could be applied to real assets and used to work out the pricing of investment alternatives based on future cash flows for the alternatives. Merton's method—named *real options analysis*—became popular among professionals as it offered more flexibility than the classical NPV methods and the options represented alternative foresight scenarios for development that could not be worked out with the NPV methods. At IAMSR the real options analysis has been applied to (i) so-called giga investments which are large enough to change the dynamics of the markets on which the investment object operates [17]; (ii) mergers and acquisitions where the strategic positions of the new, merged company offer new and different options [1]; (iii) portfolios of R&D projects where combinations of projects (some successful, others not) offer options to proceed with investments for own production, investments combined with licensing of technology or production capacity, licensing of patents, disinvestment and sales of patents, or with discontinuing the project as unsuccessful (Heikkilä [22]); (iv) portfolios of patents with options to apply/not apply for patents (national, EU, USA or global), to continue/discontinue the patents, to exclusively use the patents, to license the patents, to sell the rights to the patents and collect royalty, to sell the rights outright and to discontinue the patents; all these options have different profiles of expected revenue, risk and cash flows (Wang [30]).

In the following, we will use experience from a real world case—the strategic decision on the closing/not closing of a paper mill in the UK owned by a Finnish forest corporation. We worked with the management team during an 18 month period and followed the processes they went through and tried to support them with good analytical tools as best we could. In this way we gained a fairly good understanding of how management works with analytics tools, how they can formulate their insight with the elements of real options models, what compromises and simplifications they are ready to accept and how they understand and apply the results (cf. [21, 22]). This is now more than digital coaching for the managers, it is actually practical real world coaching.

2.1 A Paper Mill Case

The paper mill has had an unsatisfactory profitability development for a number of reasons: (i) fine paper prices have been going down for 6 years (ii) costs are going up (raw material, energy, chemicals) (iii) demand is either declining or growing slowly depending on the markets (iv) production capacity cannot be used optimally, and (v) the £/USD exchange rate is unfavourable (sales invoiced in USD, costs paid in £). The standard solution for most forest industry corporations is to try to close any old, small and not cost-effective production plants (like the UK plant); the common wisdom is that modern, large production plants will always be more profitable. The UK paper mill is producing fine paper products, it is rather aged, the paper machines were built a while ago, the raw material is not available close by, energy costs are reasonable but are increasing in the near future, key domestic markets are close by and export markets (with better sales prices) will require investments in the logistics network.

The intuitive conclusion, based on the facts, is—a sunset case and senior management should close the plant. On the other hand we have the UK trade unions, which are strong, and we have pension funds commitments for many more years which are very strict, and we have long-term power contracts which are expensive to get out of. Finally, by closing the plant we will invite competitors to fight us in the UK markets we have served for more than 50 years and which we cannot serve from other plants at any reasonable cost.

It is clear that the decision problem is more complex than standard routine formulations and that a number of factors that will decide the outcome will not be easily handled with algorithms.

There were a number of conditions which were more or less predefined. The *first* one was that no capital could/should be invested as the plant was regarded as a sunset plant. The *second* condition was that we should in fact consider five scenarios: the current production setup with only maintenance of current resources and four options to switch to setups that save costs and have an effect on production capacity used. The *third* condition is that the plant together with another unit should carry sizeable administrative costs of the sales organization which should be covered in some way (but not clear how) if the plant is closed. The *fourth* condition is the pension scheme that needs to be financed for several more years. The *fifth* condition is given by the power contracts that are also running for several more years. These specific conditions have consequences on the cost structure and the risks that various scenarios involve. It is not known if the conditions are truly non-negotiable. The management team should decide if the plant will (i) be closed as soon as possible (ii) not be closed, or (iii) be closed at some later point of time (and then at what point of time).

The first step to decide on the best decision is typically to carry out a profitability analysis. Modern profitability analysis works with methods that originate in neoclassical finance theory. The models are by nature normative and offer general decision support for the long run but may not be helpful for real life decisions in a real industry setting where data is neither well-structured nor complete. In profitability planning a good enough solution is many times both efficient, in the sense of smooth management processes, and effective, in the sense of finding the best way to act, as compared to theoretically optimal outcomes.

Economic feasibility and profitability are key factors but more issues are at stake. Management decisions will be scrutinized and questioned regardless of what the close/not close decision is going to be. The shareholders will react negatively if they find out that share value will decrease (closing a profitable plant, closing a plant which may turn profitable, or *not* closing a plant which is not profitable, or which may turn unprofitable) and the trade unions, local and regional politicians, the press, etc. will always react negatively to a decision to close a plant almost regardless of the reasons.

Only very few decisions are of the type *now-or-never*—often it is possible to postpone, modify or split up a complex decision in strategic components, which can generate important learning effects and therefore essentially reduce uncertainty. If we close a plant we lose all alternative development paths which could be possible under changing conditions. These aspects are widely known—they are part of managerial common wisdom—but they are hard to work out unless we have the analytical tools to work them out and unless we have the necessary skills to work with these tools.

The intention here is to demonstrate that in industrial cases the focus of the standard NPV based methods is too narrow; the net present value of estimated future revenues and costs gives an over-simplistic view and comparison of the decision alternatives; nevertheless, this approach is used as better tools are not readily known.

2.2 Real Options Modelling

We chose to work with real options models as our analytical tools for the paper mill. The rule we worked out, is that *we should only close the plant now if the net present value of this action is high enough to compensate for giving up the value of the option to wait*. Because the value of the option to wait vanishes right after we decide to close the plant, this loss in value is actually the opportunity cost of our decision (cf. Alcaraz [1], Borgonovo and Peccati [3], Carlsson and Fullér [6, 9], Heikkilä [22]). This is a principle based in theory but it turned out that it was well understood by the management team and the managers we worked with were interested in learning to use the real options methods. We worked out the rather advanced mathematics in a series of workshops in which we also introduced and demonstrated the software (actually Excel models) we were using—the key turned out to be that we used the management team's own data to explain the models step by step. They could identify the numbers and fit them to their own understanding of the close/no close alternatives and their consequences and the possible problem solving paths shown by the real options models. This is a good example of how coaching can be built to work in practice—managers have lots of experience, ideas, advice from various sources and influence from stakeholders that all could intervene, sometimes unintentionally, to make the decisions for them; systematic modelling will not get time and space even if it is found out that the results may be vital for the actual decision making.

The value of a real option is computed by (cf. Carlsson et al. [9, 10], Collan et al. [16], Collan [17]),

$$\mathrm{ROV} = S_0 e^{-\delta T} N(d_1) - X e^{-rT} N(d_2),$$

where

$$d_1 = \frac{\ln(S_0/X) + (r - \delta + \sigma^2/2)T}{\sigma\sqrt{T}},$$
$$d_2 = d_1 - \sigma\sqrt{T}$$

Here, S_0 denotes the present value of the expected cash flows, X stands for the nominal value of the fixed costs, r is the annualized continuously compounded rate on a safe asset, δ is the value lost over the duration of the option, σ denotes the uncertainty of the expected cash flows, and T is the time to maturity of the option (in years). The interpretation is that we have the difference between two streams of cash flow: the S_0 is the revenue flow from the plant and the X is the cost generated by the plant; both streams are continuously discounted with a chosen period of time T and the streams are assumed to show random variations, which is why we use normal distributions N. In the first stream we are uncertain about how much value δ we will lose if we postpone the decision and in the second stream we have uncertainty on the costs σ.

The function N (d) gives the probability that a random draw from a standard normal distribution will be less than d, i.e. we want to fix the normal distribution,

$$N(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d} e^{-x^2/2} dx.$$

Facing a deferrable decision, the main question that a manager primarily needs to answer is the following: *how long should we postpone the decision—up to T time periods—before (if at all) making it?* In a managerial context that is normally not decided by any algorithm but by experience, advice from people—or (in the present context) from impressions gained in negotiations.

With the model for real option valuation we can find an answer and develop the following natural decision rule for an optimal decision strategy (cf. Carlsson and Fullér [5, 8–10]).

Let us assume that we have a deferrable decision opportunity P of length L years with expected cash flows $\{cf_0, cf_1, \ldots, cf_L\}$, where cf_i is the cash inflows that the plant is expected to generate in year i ($i = 0, \ldots, L$). We note that cf_i is the anticipated net income (revenue—costs) of decision P year i. In these circumstances, if the maximum deferral time is T, we can make the decision to postpone for t' periods (which is to exercise the option at time t', $0 < t' < T$) for which the value of the option, $ROV_t{}'$ is positive and gets its maximum value; namely (cf. Carlsson and Fullér [9] for details),

$$ROV_{t'} = \max_{t=0,1,\ldots,T} ROV_t = \max_{t=0,1,\ldots,T} V_t e^{-\delta T} N(d_1) - X e^{-rT} N(d_2) > 0,$$

If we make the decision now without waiting, then we will have

$$ROV_0 = V_0 - X = \sum_{i=0}^{L} \frac{cf_i}{(1+\beta_P)^i} - X.$$

That is, this decision rule also incorporates the net present valuation of the assumed cash flows; β_P is the risk-adjusted discount rate of the decision. This is the rule for how long we can postpone the decision; this is anchored in solid economic theory which is a rational motivation for the decision. The real option model actually gives a value for the deferral which makes it possible to find the optimal deferral time. The management team gets an instrument for the decision to be made.

2.3 Real Options and Real World Decisions

Having got this far we will have to face another problem: the difference between management science modelling and what is possible in the real world case. Real options theory requires rather rich data with a good level of precision on the expected future cash flows. This is possible for financial options and the stock market (following the effective market hypothesis) for which we can use models that build on stochastic processes and which have well known mathematical properties. The data we could collect on expected future cash flows were not precise and were incomplete; the management team was rather reluctant to offer any firm estimates (for very understandable reasons, these estimates can be severely questioned with the benefit of hindsight). It turns out that we could work out the real options valuation also with imprecise and incomplete data.

Let us now work out the case that expected cash flows of the close/not close decision cannot be characterized with single numbers. With the help of possibility theory (cf. Carlsson and Fullér [7, 9] for details) we can estimate the expected incoming cash flows at each year of the project by using a trapezoidal possibility distribution of the form

$$\overline{V_i} = (s_i^L, s_i^R, \alpha_i, \beta_i), \quad i = 0, 1, \ldots, L,$$

that is, the most possible values of the expected incoming cash flows lie in the interval $[s_i^L, s_i^R]$ (which is the core of the trapezoidal fuzzy number describing the cash flows at year i of the paper mill); $(s_i^R + \beta_t)$ is the upward potential and $(s_i^L - \alpha_t)$ is the downward potential at year i, $(i = 0, 1, \ldots, L)$. In a similar manner, we can estimate the expected costs by using a trapezoidal possibility distribution of the form

$$\overline{X} = (x^L, x^R, \alpha', \beta'),$$

i.e. the most possible values of the costs lie in the interval $[x^L, x^R]$; $(x^R + \beta')$ is the upward potential and $(x^L - \alpha')$ is the downward potential (there should actually be different costs for each year, but the management team stated that they do not change much and that the trouble of estimating them does not have a good trade-off with the accuracy of the model).

By using possibility distributions we can extend the classical probabilistic decision rules for an optimal decision strategy to the case with imprecise data.

Let P be a deferrable decision opportunity with incoming cash flows and costs that are characterized by the trapezoidal possibility distributions given above. Furthermore, let us assume that the maximum deferral time of the decision is T, and the required rate of return on this project is β_P. In these circumstances, we should make the decision (exercise the real option) at time t', $0 < t' < T$, for which the value of the option, C_t' is positive and reaches its maximum value. That is,

$$\overline{FROV_{t'}} = \max_{t=0,1,\ldots,T} \overline{FROV_t} = \max_{t=0,1,\ldots,T} \overline{V_t} e^{-\delta t} N(d_1^{(t)}) - \overline{X} e^{-rt} N(d_2^{(t)}) > 0,$$

where

$$d_1^{(t)} = \frac{\ln\left(E(\overline{V_t})/E(\overline{X})\right) + (r - \delta + \sigma^2/2)t}{\sigma\sqrt{t}},$$

$$d_2^{(t)} = d_1^{(t)} - \sigma\sqrt{t} = \frac{\ln\left(E(\overline{V_t})/E(\overline{X})\right) + (r - \delta - \sigma^2/2)t}{\sigma\sqrt{t}}.$$

Here, E denotes the possibilistic mean value operator and

$$\sigma = \sigma(\overline{V_t})/E(\overline{V_t})$$

shows the annualized possibilistic variance of the aggregate expected cash flows relative to its possibilistic mean. Furthermore,

$$\overline{V_t} = \mathrm{PV}(\overline{cf_0}, \overline{cf_1}, \ldots, \overline{cf_L}; \beta_P) - \mathrm{PV}(\overline{cf_0}, \overline{cf_1}, \ldots, \overline{cf_{t-1}}; \beta_P) = \mathrm{PV}(\overline{cf_t}, \ldots, \overline{cf_L}; \beta_P) = \sum_{i=t}^{L} \frac{\overline{cf_i}}{(1+\beta_P)^i}$$

computes the present value of the aggregate (fuzzy) cash flows of the project if this has been postponed t years before being undertaken.

To find a maximizing element from the set

$$\left\{\overline{FROV_0}, \overline{FROV_1}, \ldots, \overline{FROV_T}\right\}$$

we need a method for the ordering of trapezoidal fuzzy numbers. This is one of the partially unsolved problems for fuzzy numbers as we do not have any complete model for ranking intervals (cf. Carlsson and Fullér [7, 9] for details), which is why we have to resort to various ad hoc methods to find a ranking. Basically, we can simply apply some value function to order fuzzy real option values of trapezoidal forms

$$FROV_t = (c_t^L, c_t^R, \alpha_t', \beta_t'), \quad t = 0, 1, \ldots, T.$$

$$\nu(FROV_t) = \frac{c_t^L + c_t^R}{2} + r_A \cdot \frac{\beta_t' - \alpha_t'}{6},$$

where $r_A \exists 0$ denotes the degree of risk aversion. If $r_A = 1$ then trapezoidal fuzzy numbers are compared by their pure possibilistic means (cf. Carlsson and Fullér [6]). Furthermore, in the case $r_A = 0$, we are risk neutral and fuzzy real option values are compared by the centre of their cores.

Thus we can work out the best time for making a close/not close decision on the paper mill also with imprecise and incomplete data, that is—if we can work out the mathematics. If this is not the case and our knowledge and expertise is to be found in business operations (the mathematics is basically known but details are distant memories) we would benefit from *coaching*. We have worked with managers in a number of real option cases (cf. [8]) and have found out that they could benefit from the following types of coaches, which now are at the analytics level and in the domain of models and algorithms:

- *Coach 1*: collects and guides estimates of imprecise incoming cash flows and costs, shows FROV with guides to meaning, and shows the T periods for optimal postponement of the decision with guides to explanation (net present value of action vs. option value of postponement)
- *Coach 2*: collects, guides market estimates for r (compound rate), δ (value lost over the option) and σ (uncertainty of expected cash flows); guides estimates of α' and β', the downward and upward potentials
- *Coach 3*: guides to explanations of FROV, possibilistic mean, possibilistic variance, present value of aggregate fuzzy cash flows
- *Coach 4*: guides to explanations of ranking of FROV (ranking of fuzzy numbers)
- *Coach 5*: works out trade off variations between net present value of action vs. option value of postponement with variations of parameters

The immediate reaction is of course that this coaching could be carried out by a senior expert with knowledge and experience of real options and the key factors in decisions to close an old paper mill.

Experience from real life cases shows that this is both expensive and impractical and that online digital coaching would be both less expensive and more practical; the remaining problems are to find some effective methods to build and implement digital coaches. In the actual case we were able to use Excel to implement the FROV models; Coach 1 would guide estimates of the cash flows for the Excel model and explain the optimal solutions; Coaches 2–4 would guide parameter estimates and explain the real options and fuzzy theory parts of the model; Coach 5 would guide simulations with the models and offer explanations of the results.

In the next section we will find out how digital coaches could be constructed to offer the described functionality without supervision of personal coaches.

3 Digital Coaching

The digital coaching systems got started a few years ago as an answer to the demand on human operators to master advanced automated systems that are used to monitor and control often complex and very large industrial process systems. Digital coaching will work on and with data, information and knowledge that is collected from digital devices, instruments, tools, monitoring systems, sensor systems, software systems, data and knowledge bases, data warehouses, etc. and then processed to be usable for the digital systems that will guide and support users. The processing is done with digital fusion which operates in three phases: data, information and knowledge fusion.

Data fusion is the first step; a function that combines multiple tuples into one is called a fusion function and the standard, rather simple operation is a fusion of data attributes. The traditional way is to define some ordering relation a priori and then to keep it updated for continuous use; maintenance is challenging for big, fast data which is why we want automatic support in modern industrial applications. A better than the traditional way is to construct order relations automatically when data attributes to be fused are inspected.

Data fusion builds data sets (or families of data sets). The next step is to extract process *information* from the (often big) data sets with analytics methods such as data mining, statistical analysis, machine learning, computational intelligence, visualization, etc. The process is continued with analytical techniques to fuse sets of information to more meaningful summary information—*information fusion* (cf. Carlsson et al. [11, 12, 14, 15]). Some early research results show that information fusion reduces the uncertainty in (social) big data by extracting key (valid, relevant) factors, cleaning out outliers, high-lighting illogical assumptions, etc. (cf. Morente-Molinera et al. [28, 29]).

Knowledge fusion builds on first data fusion, then information fusion. Knowledge fusion applies taxonomies or ontologies—in the D2I program (funded by Tekes 340/12) fuzzy ontology was developed and used to detect, identify and deal with recurring problems in pump valve packages (Carlsson et al. [13–15]). Automated knowledge builds on natural (or near-natural) language processing, information extraction with analytics tools, information integration (or federation), computational intelligence (soft computing, evolutionary computation, swarm intelligence, intelligent agents, etc.) (cf. Morente-Molinera et al. [28, 29]).

With the proposed basis in digital fusion we can then look for theory and technology constructs that could be used for the digital coaches.

One of the approaches to *virtual coaches* builds on the emerging technology of embodied conversational agents (ECA's). ECA's are animated virtual characters, displayed on a computer or a mobile device screen. ECA's play the roles of teachers, mentors, advisors, social companions, and, increasingly, of virtual coaches (cf. Hudlicka [23]). The ECA's engage in natural interaction with humans through dialogue and non-verbal expression which requires minimal or no training; we will probably not be aided by animated virtual characters but the core agent constructs may be useful for the digital coaching when working on data and information fusion material. With some quick sketching we could work with the following setup and functionality: $Coach_i$ will (i) collect incoming cash flows (*data fusion*); (ii) carry out estimates of incoming cash flows (*information fusion*), and (iii) build explanations of NPV of action versus option value of postponement (*knowledge fusion*).

The *virtual trainer* systems are becoming popular as supporting services to fitness and wellness applications; they are typically identified as three classes (i) smart phone applications (ii) sensor devices, and (iii) image processing devices. Sensor data and images are collected and processed through data fusion, which can be a low level implementation that builds on fast, efficient process monitoring. In our present context this could be used for working out the parameters needed for the Markov processes. The "trainer" gives feedback on the progress of the exercise and offers summary post exercise data for learning and for motivation to keep up the exercises. This basic functionality appears to be generic and we should bear it in mind for the *digital coaches*.

3.1 Coaching with Markov Decision Processes

Fern et al. [19] work out a theory base for personalised AI systems that work as personal assistants to support human users with tasks they do not fully know how to carry out. This type of technology has gained much attention in the last 10 years because of the growing use of automated systems with intelligent functions. Fern et al. [19] work out a model where the assistant (an AI system) observes the user (represented as a goal-oriented agent) and must select assistive actions from a closed set of actions in order to best help the user achieve his goals.

The context studied is a physical environment in which the assistant helps a user to navigate, follows up on the progress, adjusts the behaviour towards some chosen goal and continues with sequential adjustments until the user is satisfied with the goal attainment. The functionality is close to the one we have in mind for the Coaches 1–5. The interesting thing is that this functionality builds on a decision-theoretic model, which is worked out with partially observable Markov decision processes (POMDPs). Fern et al. [19] work out variations of these Markov processes to get a formal basis for designing intelligent assistants. A specific case is the hidden goal Markov decision processes (HGMDP) that are designed to cover the application environment and the user's policy and hidden goals. The HGMDP is a tuple $<S, G, A, A', T, R, \pi, I_S, I_G>$ where S is a set of states, G is a finite set of possible user goals, A is a set of user actions, A′ is the corresponding set of assistant actions, T is a transition function that decides the transition from s to s' (element of S) after the user takes action a (element of A) towards a goal g (element of G); R is a reward function for both the user and the assistant, π is the user's (optimal) policy mapping to the context, and I_S the initial and I_G the goal states. Markov processes are generic constructs that can be used to describe complex processes in a fairly compact form. The digital coaching, we want to build, appears not to be neither a POMDP nor an HGMDP because (A, A′) are not stochastic but decided by the ROV model, the transition function T is a function of the ROV and the R could be a (fuzzy) distance function.

Nevertheless, there are some useful constructs that build on the Fern et al. [19] theoretical framework. The first is a special case where the assistant's policy is deterministic for each specific goal. This opens up for the use of an optimal trajectory tree (OTT) where the nodes represent the states of the MDP reached by the prefixes of optimal action sequences for different goals starting from the initial state. Each node in the tree represents a state and a set of goals for which it is on the optimal path from the initial state. The size of the optimal trajectory tree, which we need to be reasonably compact for computational purposes, is bounded by the number of goals times the maximum length of any trajectory, which is at most the size of the state space in deterministic domains. This gives some hints at what constructs to look for when trying to work out the Coaches 1–2 that decide the parameter values for the ROV model.

Another interesting result is the approach to solve the problem of selecting an assistive action. For an HGMDP Fern et al. [19] work out a combination of bounded look-ahead search and myopic heuristic computations (selecting an action that has the highest probability of being accepted). By increasing the amount of look-ahead search the actions returned will be closer to optimal at the cost of more computation; for many HGMDPs the useful assistant actions can be computed with relatively little or no search.

We will need some constructs to learn (e.g.) an HGMDP while interacting with the context, i.e. the assistant should follow how the user interacts with the context and learn the user's policy and goal distributions. These constructs would be useful when the assistant is called upon many times for the same construct (quite often at irregular intervals) by the same user; a further extension would be for the assistant to start from basic constructs obtained with one user and then to learn another user's policy and goal distributions. The classical approach is to use Maximum Likelihood estimates of the user's policy distributions from continuous follow-ups and combine that with estimates of the goal attainment (e.g. as fuzzy distances). Another approach that Fern et al. [19] propose is to use an MDP model of the context and bootstrap the learning of the user policy. This would be useful if the user is near optimal in his policy choices and will likely select actions that are near-optimal for a selected goal and an actual context.

3.2 Coaching with Virtual Environments

Fricoteaux et al. [20] work out the use of virtual environments for fluvial navigation; these environments offer training in easily modifiable environmental conditions (wind, current, etc.), which have impacts on the behaviour of a ship; fluvial navigation would in our context be applied to rapidly and randomly changing market conditions. The main difficulty in fluvial navigation is to anticipate manoeuvres and the variability of the conditions of the environment. It is interesting to note that the formal framework to represent and support the decision-making system builds on classical Dempster-Shafer theory in order to take account of uncertainty. Unlike the theory of probability, the DS-theory allows for explicit modelling of ignorance. This can be combined with directed graphs to represent influences between variables; if the inference is probabilistic, Bayesian networks (BNs) can be used; with belief functions there are evidential networks with conditional belief functions (ENCs). Then in turn ENCs have been generalised by evidential networks with conditional belief functions (DEVNs), etc.—thus there appears to be constructs available to apply to the building of digital coaches. The remaining challenge appears to get it done.

Bloksmal and Struik [2] work out a program for coaching farmers using human health as a metaphor for farm health, which helps both them and the farmers to gain an understanding of the issues that are crucial for improving the processes and the productivity of a farm. The coach and the farmer together work out the course of life of the farm, they learn from what has happened in the history of the farm and translate images of possible futures into the current state of the farm. The shorter-term issues, making work more rewarding, improving work efficiency and effectiveness, farm productivity and profit are worked out in this context and longer-term issues such as organisational and spatial redesign of the farm are worked on against this background. The coach operates like a physician—“alternatingly observing the diseased part and the whole being of an ill person”—and by referring to this metaphor, opens up similar mechanisms for what may be wrong with the farm. If we do this skilfully, it will show to what extent the farm resembles a living and healthy entity and the farmer will get new ideas on how to improve “the living farm organism”; the use of metaphors takes out the blame from the narrative, the farmer will not feel that the coach blames him for having done something wrong.

Bloksmal and Struik [2] show that the process to find and describe the identity and key processes of a farm is not easy; they use a narrative method—the coach listens to discover the drama behind the facts. This approach could work for our digital coaching of real options modelling and planning—i.e. the real options models could be good representations of the “drama behind the facts” and would show the key relations to decide the optimal interval for the postponement of a decision with explanations that a manger could understand and make use of.

4 Real Options Support for Decisions on a Paper Mill

The fuzzy real options models were used to make a decision on the old paper mill in the United Kingdom. The decision had two parts (i) to decide if the paper mill should be closed or not, and if (i) is positive, then (ii) the optimal point in time when to close the mill (actually the optimal number of periods to postpone the closing from present time, cf. [21, 22] for details).

The implementation of the real options theory for a practical case entails simplifications, mostly due to lack of either data, information or knowledge, or in order to be able to build the actual real options model. This is a typical trade-off we have to make in order to use a theory, but if we simplify too much we will stray too far from the core of the theory which makes the results questionable (cf. [26, 27]).

We built scenarios in order to outline possible developments with the information and knowledge we could get about the plant, the product lines, the customers, the competition, the markets and the country context offered by the UK. This produced data for us to work on but data came from different sources and related to different periods, which gave somewhat inconsistent and heterogeneous scenarios. As we could not be very precise, the managers argued that simplifications would not add much to the uncertainty of the views of the future worlds.

Each scenario assumes a match between sales and production, which is a simplification; in reality there are significant, stochastic variations in sales which cannot be matched by the production. Since the planning assumed no capital investments, there will be no costs in switching between the scenarios (which is another simplification). We worked out some exercises on the possibilities to switch in the future as (real) options for senior management; the opportunity to switch to another scenario is a call option. The option values build on the estimates of future cash flows, which are the basis for the upward/downward potentials.

Senior management (reluctantly) adopted the view that options can exist and that there is a not-to-decide-today possibility for the close/not-close decision. The motives to include options into the decision process built on the following logic:

- New information changes the decision situation (“good or bad news”)
- Consequently, new information has a value and it increases the flexibility of the management decisions
- The value of the new information can be analysed to enable the management to make better informed decisions

In the discussion, we were able to show that companies fail to invest in valuable projects because planners overlook and leave out the options embedded in a project from the profitability analysis. The real options approach shows the importance of timing as the real option value is the opportunity cost of the decision to wait in contrast with the decision to act immediately.

We were then able to give the following practical description of how to form the option value:

$$\begin{aligned} \textit{Option value} = & \textit{ Discounted cash flow} \\ & \times \textit{Value of uncertainty } (\textit{usually standard deviation}) - \textit{Investment} \\ & \times \textit{Risk free interest} \end{aligned}$$

If we compare this sketch with the decision to close/not-close the production plant with the FROV models we introduced we cannot avoid the conclusion that things are much simplified. The substance of the decision is not the same, which we could not describe very well without getting into the stochastic processes at the core of the real options theory—and this was beyond what we could reasonably expect the managers to follow. Then we run into another, more serious problem—senior management will distrust results of an analysis they cannot evaluate and verify with numbers they recognize or can verify as "about right".

We found out—somewhat unexpectedly—that we could build the fuzzy real options models using Excel; this solved part of the problem, as the managers we worked with were semi-professional Excel users and could figure out how the models work by experimenting with Excel. Another part of the solution was that we could include actual numbers from the plant in the scenarios and the managers could judge from the outcome that they were "about right", "reasonable" and "verifiable". Then we (of course) added NPV to the Excel models to allow the managers to test their intuitive understanding with verifications carried out through familiar NPV calculations.

We found out that the added functionality of *Coach 1* and *Coach 2* would have simplified the implementation and use of the real options models; senior decision makers want to *know* how the key factors work and interact, what the outcome is going to be and then what consequences can be expected from the outcome (Fig. 1).

In the same way, the combined use of *Coach 1* and *Coach 5* will be helpful to work out and explain the scenario combinations over time; *Coach 5* will show the trade-off between the net present value of action now (*closing the paper mill now*) and the option value of postponing the decision (*option value = 0 if action now*) (Fig. 2).

The detailed reports are for the managers so that they could check how reasonable and verifiable the outcome could be (the numbers are modified and the timeframe is altered as the data was highly confidential). *Coach 3* explains what the possibilistic (fuzzy) values are, explains the FROV and shows and explains the assumptions underlying the NPV of cash flows (which are fuzzy numbers).

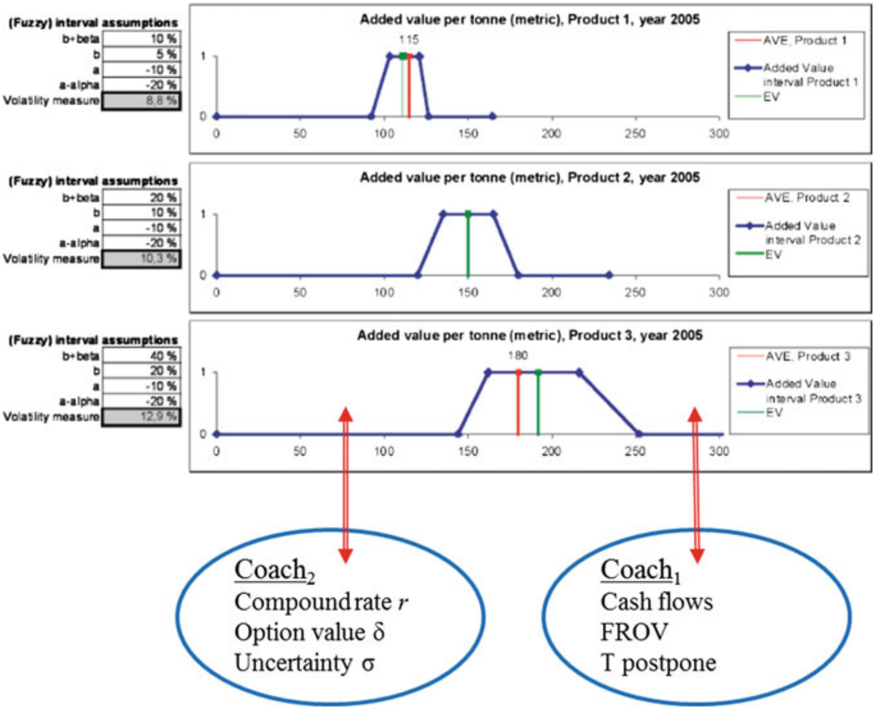


Fig. 1 Scenario alternatives included estimated cash flows, compound rate, option value of a postponement, uncertainty of incoming cash flows, calculated FROV and optimal postponement T

Coach 4 works out and explains the ranking of the FROV values, which actually is a ranking of fuzzy numbers (cf. [12, 13]). The functionality of the last two coaches show issues that were questioned and tested in order to get at the core of the models and the results (Fig. 3).

5 Summary and Future Scenarios

The theory framework built by and for the classical Operations Research and Management Science since the early 1950s is now in the process of being transformed by (Business) Analytics, which is getting the attention of major corporations and senior management. The research groups working at IAMSR have been part of this process and have specialised in working out and using analytics methods, which implement soft computing theory and algorithms (cf. [17, 22, 26, 27]). This has turned out to be very effective and useful for planning, problem solving and decision making in "big data" environments; the "big data" is one of the challenges

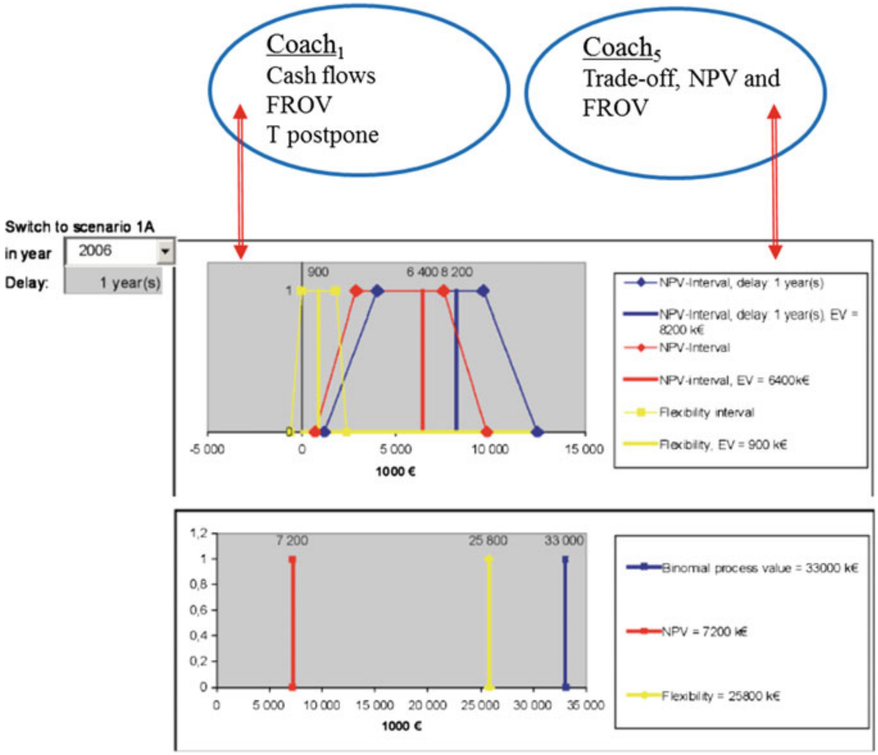
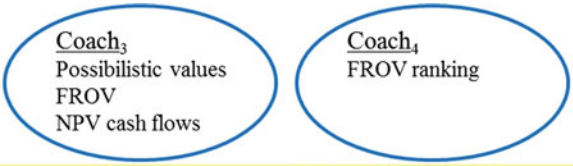


Fig. 2 Outcome of scenario alternatives worked out as FROV (flexibility), NPV and a binomial process value

of the modern digital economy and for which we propose that fuzzy sets would offer instruments for breakthrough research.

Analytics adds value to management; it promotes data-driven and analytical decision-making and "reinvented" fact-driven management. Analytics builds on recent software improvements in information systems that has made data, information and knowledge available in real time in ways that were not possible for managers only a few years ago (Davenport and Harris, [18]).

In the context of the digital economy common wisdom finds that real-time management is a necessity as operations should be planned and carried out in a fast changing and complex environment where careful and thoughtful management will be bypassed by fast, innovative approaches (which may turn out to be of inferior quality, but have then already established sustainable competitive positions). Real-time management is challenged by "big data" and the necessity for fast processing using advanced analytics methods. The advanced analytics would require postdoc-qualified managers—these are rather scarce in senior management positions. Thus there will be a need to reinstate "coaching" functions with the advanced analytics methods to tell/explain to the users what can/should be done, how it



	2004	2005	2006	2007
Present value at delay		7,174,624		6,494,629
Present value at delay, Support up		9,834,912		14,886,532
Present value at delay, Core up		7,552,125		11,824,291
Present value at delay, Core down		2,986,552		5,699,809
Present value at delay, Support down		703,765		2,637,568
Present value at delay, Fuzzy EV		6,410,732		10,293,171
Present value at delay, St. Dev.		2,345,340		3,146,154
Present value at delay, St. Dev. %		36.6%		30.6%
NPV at present year, 2005	Flexibility			
Delay value without flexibility	-1,283,804	7,174,624		5,890,820
Delay value with flexibility, Support Up	3,667,612	9,834,912		13,502,524
Delay value with flexibility, Core Up	3,172,855	7,552,125		10,724,981
Delay value with flexibility, Core Down	2,183,343	2,986,552		5,169,895
Delay value with flexibility, Support Down	1,688,587	703,765		2,392,352
Delay value with flexibility, Fuzzy EV	2,925,477	6,410,732		9,336,209
Delay value with flexibility, St. Dev.	508,314	2,345,340		2,853,654
Delay value with flexibility, St. Dev. %	17.4%	36.6%		30.6%
Delay	2			

Fig. 3 Numerical outcome reports on the scenario alternatives

should be carried out, what the results are and what they mean, and how they should be applied (with explanations of what could/should not be done).

We have worked out a context and a scenario for digital coaching with the help of real options modelling. Black and Scholes introduced the options theory to decide the price for options on financial assets; Merton proved that similar options modelling could be applied to real assets and used to work out the pricing of investment alternatives based on future cash flows for the alternatives. Merton's method—named real options analysis—became popular among professionals as it offered more flexibility than the classical NPV methods.

We illustrated the development and use of fuzzy real options models with the case of closing (or not closing or closing later) an old paper mill in the UK. The rule we worked out was that *we should only close the plant now if the net present value of this action is high enough to compensate for giving up the value of the option to wait*. Because the value of the option to wait vanishes right after we decide to close the plant, this loss in value is actually the opportunity cost of our decision. This is a principle based in theory but it turned out that the management team could well accept the principle—and then wanted to find out how to use it.

We discussed the need for coaching, but have found through experience from real life cases that it would be both expensive and impractical to try to find and use experienced human coaches. The alternative is online digital coaching and we only need to find some effective methods to build and implement digital coaches. We found out that there are not many useful approaches offered in the literature, the closest we could get was an elaborate framework built around Markov decision processes, which was not actually up to the task. In order to outline what the coaches should do we worked out the processes in terms of the models we developed for the old paper mill in the UK. We were able to use Excel to implement the FROV models in the actual case, which was a bit of a surprise. We could then describe that *Coach 1* would guide estimates of the cash flows for the Excel model and explain the optimal solutions; *Coaches 2–4* would guide parameter estimates and explain the real options and fuzzy theory parts of the model; *Coach 5* would guide simulations with the models and explain the results. We will work out some possible theoretical frameworks for the digital coaches in a coming series of papers.

We will conclude with some arguments for why it will make sense to work with analytical models even in a context, which is not recognized as a domain for analytics. Kahneman [24] relates the case of Orley Ashenfelter, a Princeton economist and wine lover, who wanted to find a way to predict the future value of fine Bordeaux wines from information available in the year they are made (cf. [24]). He was of course well aware that he was stepping on the sensitive toes of world-renowned experts who claimed that they could predict the value development for individual wines over years to come and also in which year the wines will reach their peak quality and highest price. The experts built their judgement on tasting the wines and decades of experience of and insight in the wine markets; Ashenfelter built his predictions on regression analysis and an effective use of statistics tools—he had no possibility to actually taste the wines. Ashenfelter collected statistics on London auction prices for select mature red Bordeaux wines 1990–91 (sold in lots of a dozen bottles); mature red Bordeaux were defined as vintages 1960–1969 and the wines selected came from six Bordeaux chateaux which are large producers with a reputation to have produced high quality wine for decades—or centuries in some cases. Ashenfelter found out that the quality of the Bourdeaux wines is decided by (i) the age of the vintage (ii) the average temperature over the growing season (April–September) (iii) the amount of rain in September and August (less rain gives better wine), and (iv) the amount of rain preceding the vintage (October–March). These four factors are all measurable and built on published and easily verifiable facts; Aschenfelter collected data on the vintages 1952–1980 and built a regression model with the four factors which turned out to explain about 80% of the variation in the average price of Bordeaux wine vintages. *His point is that the future quality of Bordeaux wines can be worked out without tasting the wines or introducing any kind of subjective judgements*. He used his models to predict the price development for new vintages of Bordeaux (the correlation between prediction and actual prices is above 0.90) which he has shared with a crowd of followers that are investing in promising, good vintages. Through his models he has also found a few very good vintages that are under-priced in the market and which he and his friends have bought and much enjoyed.

First we can notice that Aschenfelter made sure that he had observations on large selections of wine over 10 years from six large chateaux—but only in Bordeaux in order to reduce the number of external factors that influence the wine production but are not relevant for the key issues of his study. Second, his models forecast future prices (years, and even decades into the future) more accurately than the current market prices of young wines do that build on expert estimates;

this challenges economics theory that claims that market prices (in effective markets) will reflect all information on the products. Third, experts make judgements that are inferior to algorithms; Kahneman (cf. [24] p. 224) argues that some reasons for this is that experts try to be clever, to think outside the box and to work with (too) complex combinations of features to make their predictions. Complexity may work in specific cases but will reduce validity in most cases. There is a *second point* to be made—*analytics, when the proper methods are developed and used, will give insight that intuition and experience will not be able to produce*. This is a lesson learned for the digital economy where it is claimed that the dynamics of the market and the need to make (almost) real time decisions in order to stay competitive makes it necessary to forego analytics and rely on the intuition and experience of visionary managers and executives.

References

1. Alcaraz Garcia, F.: Real options, default risk and soft applications, 82 pp. TUCS Dissertations (2006)
2. Bloksmal, J.R., Struik, P.C.: Coaching the process of designing a farm: using the healthy human as a metaphor for farm health. NJAS **54**(4), 413–429 (2007)
3. Borgonovo, E., Peccati, L.: Sensitivity analysis in investment project evaluation. Int. J. Prod. Econ. **90**, 17–25 (2004)
4. Carlsson, C.: On the relevance of fuzzy sets in management science methodology. In: Zimmermann, H.J., Zadeh, L.A., Gaines, B.R. (eds.) Fuzzy Sets and Decision Analysis, TIMS Studies in Management Sciences, vol. 20, pp 11–28. North-Holland, Amsterdam (1984)
5. Carlsson, C., Fullér, R.: Capital budgeting problems with fuzzy cash flows. Mathw. Soft Comput. **6**, 81–89 (1999)
6. Carlsson, C., Fullér, R.: On optimal investment timing with fuzzy real options. In: Proceedings of the EUROFUSE 2001 Workshop on Preference Modeling and Applications, pp. 235–239 (2001)
7. Carlsson, C., Fullér, R.: On possibilistic mean value and variance of fuzzy numbers. Fuzzy Sets Syst. **122**, 315–326 (2001)
8. Carlsson, C., Fullér, R.: Fuzzy Reasoning in Decision Making and Optimization. Springer, Berlin, Heidelberg (2002)
9. Carlsson, C., Fullér, R.: A fuzzy approach to real option valuation. Fuzzy Sets Syst. **139**, 297–312 (2003)
10. Carlsson, C., Fullér, R., Majlender, P.: A fuzzy real options model for R&D project evaluation. In: Proceedings of the 11th IFSA World Congress (2005)
11. Carlsson, C.: Soft computing in analytics: handling imprecision and uncertainty in strategic decisions. Fuzzy Econ. Rev. **XVII**(2), 3–21 (2012)
12. Carlsson, C., Heikkilä, M., Mezei, J.: Possibilistic bayes modelling for predictive analytics. In: Proceedings of 15th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, pp. 15–20 (2014)
13. Carlsson, C., Mezei, J., Wikström, R.: Aggregating linguistic expert knowledge in type-2 fuzzy ontologies. Appl. Soft Comput. **35**, 911–920 (2015)
14. Carlsson, C., Heikkilä, M., Mezei, J.: Fuzzy entropy used for predictive analytics. In: Kahraman, C. (ed.) Fuzzy Sets in its 50th year. New Developments, Directions and Challenges, Studies in Fuzziness, vol. 341, pp. 23. Springer (2016)

15. Carlsson, C.: Imprecision and uncertainty in management—the possibilities of fuzzy sets and soft computing. NOEMA XV Rom. Acad. Sci. 89–114 (2016)
16. Collan, M., Carlsson, C., Majlender, P.: Fuzzy black and scholes real options pricing. J. Decis. Syst. **12**, 391–416 (2003)
17. Collan, M: Giga-investments: modeling the valuation of very large industrial real investments, 57 pp. TUCS Dissertations, Turku (2004)
18. Davenport, T.H., Harris, J.G.: Competing on Analytics: The New Science of Winning. Harvard Business School Press, Boston (2007)
19. Fern, A., Natarajan, S., Judah, K., Tadepalli, P.: A decision-theoretic model of assistance. J. Artif. Intell. Res. **49**, 71–104 (2014)
20. Fricoteaux, l., Thouvenin, I., Mestre, D.: GULLIVER: A decision-making system based on user observation for an adaptive training in informed virtual environments. Eng. Appl. Artif. Intell. **33**, 47–57 (2014)
21. Heikkilä, M., Carlsson, C.: A fuzzy real options model for (not) closing a production plant: an application to forest industry in finland. In: Proceedings of the 12th Annual International Conference on Real Options, Rio de Janeiro (2008)
22. Heikkilä, M.: R&D Investment Decisions with Real Options—Profitability and Decision Support, Åbo Akademi University Press, Åbo (2009)
23. Hudlicka, E.: Virtual training and coaching of health behavior: example from mindfulness meditation training. Patient Educ. Couns. **92**(2013), 160–166 (2013)
24. Kahneman, D.: Thinking Fast and Slow. Penguin Books, London (2011)
25. Liberatore, M., Luo, W.: INFORMS and the analytics movement: the view of the membership. Interfaces **41(**6), 578–589 (2011)
26. Majlender, P.: A normative approach to possibility theory and soft decision support, 54 pp. TUCS Dissertations, Turku (2004)
27. Mezei, J.: A quantitative view on fuzzy numbers, 142 pp. TUCS Dissertations (2011). Turku
28. Morente-Molinera, J.A., Wikström, R., Carlsson, C., Viedma-Herrera, E.: A linguistic mobile decision support system based on fuzzy ontology to facilitate knowledge mobilization. Decis. Support Syst. **81** (2016)
29. Morente-Molinera, J.A., Mezei, J., Carlsson, C., Viedma-Herrera, E.: Improving supervised learning classification methods using multi-granular linguistic modelling and fuzzy entropy. Trans. Fuzzy Syst. (2016)
30. Wang, X.: Fuzzy real options analysis in patent related decision making and patent valuation. Åbo Akademi Dissertations, Åbo (2015)
31. Zadeh, L.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)

An Analysis of Decision Criteria for the Selection of Military Training Aircrafts

Juan M. Sánchez-Lozano, M.A. Socorro García-Cascales and María T. Lamata

Abstract The Spanish Minister of Defense needs to replace the current military training aircrafts by other models to meet current training needs in the Spanish Air Force Academy. In order to know the main features that the candidate aircrafts should have, there is a need to take into account the knowledge and experience of experts in this specific field, such as trained test pilots and flight instructors. In this way, it will be possible to recognize the main technical criteria to consider. This study shows a case study that allowed obtaining not only the preferences of an expert's group, but also the importance of the considered criteria. Given that the criteria information provided by the experts has different nature, with qualitative criteria (human factors, flying and handling qualities, etc.) coexisting with quantitative criteria (service ceiling, stalling speed, endurance, etc.), the joint use of linguistic labels and numerical values is needed. Therefore, a survey focused on the fuzzy AHP (Analytic Hierarchy Process) methodology is proposed to extract the knowledge from the experts group and finally obtain a unique set of weights for the criteria.

J.M. Sánchez-Lozano (✉)
Centro Universitario de la Defensa. Academia General del Aire.
Universidad Politécnica de Cartagena, Murcia, Spain
e-mail: juanmi.sanchez@cud.upct.es

M.A.S. García-Cascales
Dpto. de Electrónica, Tecnología de Computadoras y Proyectos,
Universidad Politécnica de Cartagena, Murcia, Spain
e-mail: socorro.garcia@upct.es

M.T. Lamata
Dpto. Ciencias de la Computación e Inteligencia Artificial,
Universidad de Granada, Granada, Spain
e-mail: mtl@decsai.ugr.es

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_10

1 Introduction

Nowadays, the Air Force Academy uses two training aircrafts: the model ENAER T.35C Tamiz for elementary basic education and the model CASA C-101 Aviojet for advanced basic education. These aircraft have been operating from the 1980s so that, as a result of the continuous advancement of aviation technology and the high number of flight hours that they have seen, in the near future it will be necessary to replace them by other models to meet current training needs.

From the point of view of training aircrafts and, as occurs in the subsequent stages of design [1, 2], decision-making is an intellectual activity which is necessary and essential to face. Before taking any decisions, facts, knowledge and experience should be gathered to assess the context of the problem. In this type of decision-making process, a large number of essential criteria is involved. To resolve them, it is therefore advisable to employ tools such as Multi-Criteria Decision Making (MCDM), a process whose use is widespread today, not only in the military field [3–5], but also in many research fields [6–8].

In addition, when selecting the best training aircraft a number of criteria of different nature should be taken into account, such as quantitative criteria (service ceiling, stall speed, fuel range, etc.), and qualitative criteria (cockpit ergonomics, feelings of instructor, etc.). In order to model and evaluate the latter type of criteria, fuzzy logic techniques [9–13] are a good alternative, not only to operate in an isolated way but also combined with pseudo-Delphi techniques and MCDM methods (AHP [14]; TOPSIS [15]; ELECTRE [16], etc.). Although some of the aforementioned multi-criteria methods are able to apply fuzzy logic and evaluate the potential alternatives, the AHP methodology also allows obtaining the weight of the criteria. That is the main reason why in this study case a pseudo-Delphi technique has been combined with fuzzy AHP methodology.

From the point of view of the Spanish Air Force, the Air Staff and the Logistics Support Command of this force are the main decision-makers. Nevertheless, it is advisable to make a preliminary assessment taking into account the most significant technical criteria which also reflect the experience of important expert groups such as trained test pilots and flight instructors of the Spanish Air Force.

Therefore, our aim here is to determine the relative importance of the main technical criteria and then, to transform such importance into weights that should be later used in a MCDM scheme. The problem will be solved using the AHP methodology to obtain the weights of the criteria that influence the decision. Furthermore, given that the criteria are both qualitative and quantitative, both methods will be combined with fuzzy logic through the design and development of a survey to experts in the field of military training aircraft.

This chapter is divided into four sections: Sect. 2 will define the criteria that influence the decision-making, in Sect. 3 the fundamentals of fuzzy sets and the AHP methodology will be described. Section 4 will explain the way in which the weights of the considered criteria are obtained and the results. The final section will detail the conclusions of this study.

2 Decision Criteria for Evaluating Military Training Aircraft

The mission of the Spanish Air Force Academy is to train future officers of the Spanish Air Force by providing them with academic, military and aeronautical teaching. Although today there are many aircraft, they are usually classified according to their use [17]. In the case of the Spanish Air Force Academy, these aircrafts should have specific features that allow the future officers to carry out their basic and advanced education [18]. Due to that, it is highly advisable to identify the main technical criteria that influence the decision; these data have been obtained from [17, 19–22] and an advisory group composed by instructors and flight personnel of the Air Force Academy. The chosen criteria are the following:

- C_1: Service ceiling (ft), the highest operating altitude at which the maximum achievable rate of climb is 100 ft/min and the aircraft can bear the atmosphere and operate efficiently.
- C_2: Cruising speed (kt), the constant and uniform speed in which an aircraft is able to fly with normal conditions of pressure and temperature.
- C_3: Stalling speed (kt), the minimum speed in which the wings maintain lift at flameout.
- C_4: Endurance (minutes), the maximum time in which an aircraft can remain in the air until all fuel has expired.
- C_5: Positive Limit Load Factor (+ G), the maximum value of positive acceleration forces which can withstand the airframe.
- C_6: Negative Limit Load Factor (− G), the minimum value of positive acceleration forces which it can withstand the airframe.
- C_7: Take-off distance (ft), the minimum distance required by the aircraft to accelerate along the runway until it reaches a speed at which it can generate sufficient aerodynamic lift to overcome its weight (in standard sea level conditions).
- C_8: Landing distance (ft), the minimum distance required by the airplane to land (in standard sea level conditions).
- C_9: Human factors: the comfort conditions inside the cockpit (beginner pilot and instructor positions)
- C_{10}: Flying and handling qualities, confidence that the instructor or beginner pilot on the plane to perform complex training exercises.
- C_{11}: Security systems, devices of the aircraft for responding face with setbacks or unexpected situations (ejection systems, sensors, etc.)
- C_{12}: Maneuvering Capability, software tools capable of being configured and adapted to several models of education (elementary and advanced stage)

In order to determine the relative importance of these technical criteria, we have access to a group of experts (trained test pilots and flight instructors of the Spanish Air Force) who will answer a survey based on the application of the methodology described.

3 Methodology

Fuzzy Sets

The fuzzy set theory, introduced by Zadeh [9] to deal with vague, imprecise and uncertain problems has been used as a modelling tool for complex systems that can be controlled by humans but are hard to define precisely. Examples of fuzzy sets are classes of objects (entities) characterized by such adjectives as large, small, serious, simple, approximate, etc. The main reason for this is that in a real world, there are not crisp or real boundaries which separate those objects which belong to the classes in question from those which do not [23]. A collection of objects (universe of discourse) X has a fuzzy set A described by a membership function f_A with values in the interval [0,1] [24].

In this chapter, we only make reference to the operations on a triangular membership function through the fuzzy number sets that will be used in the study case. The basic theory regarding Triangular Fuzzy Numbers (TFN) is described in detail in [9]. Herein, we only make reference to the operations on fuzzy sets that we will use in the application.

Definition 1.- A_1 and A_2 are two TFN defined by the triplets (a_1, b_1, c_1) and (a_2, b_2, c_2), respectively. For this case, the necessary arithmetic operations with positive fuzzy numbers are:

(a) Addition:

$$A_1 \oplus A_2 = [a_1 + a_2, b_1 + b_2, c_1 + c_2] \tag{1}$$

(b) Subtraction:

$$A_1 \Theta A_2 = A_1 + (-A_2) = [a_1 - c_2, b_1 - b_2, c_1 - a_2] \tag{2}$$

(c) Multiplication:

$$A_1 \otimes A_2 = [a_1 \times a_2, b_1 \times b_2, c_1 \times c_2] \tag{3}$$

(d) Division:

$$A_1 \phi A_2 = [(a_1, b_1, c_1) \cdot (1/c_2, 1/b_2, 1/a_2)] \tag{4}$$

When $0 \neq [a_2, b_2, c_2]$

(e) Scalar Multiplication:

$$k \circ T_1 = (k \circ a_1, k \circ b_1, k \circ c_1) \tag{5}$$

(f) Root:

$$T_1^{1/2} = \left[a_1^{1/2},\ b_1^{1/2},\ b_1^{1/2}\right] \tag{6}$$

Analytic Hierarchy Process (AHP)

The AHP methodology, proposed by Saaty [14], has been accepted by the international scientific community as a robust and flexible MCDM tool to deal with complex decision problems. Basically, AHP has three underlying concepts:

- Structuring the complex decision as a hierarchy of goal, criteria and alternatives.
- Pair-wise comparison of elements at each level of the hierarchy with respect to each criterion on the preceding level, and finally
- Vertically synthesizing the judgements over the different levels of the hierarchy.

AHP attempts to estimate the impact of each one of the alternatives on the overall objective of the hierarchy. In this case, we shall only apply the method to obtain the criteria weights.

We assume that the quantified judgements provided by the decision-maker on pairs of criteria (C_i, C_j) are contained in an n x n matrix as follows:

$$C = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_m \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \cdots & C_n \end{array} \\ \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix} \end{array}$$

For instance, the c_{12} value represents an approximation of the relative importance of C_1 to C_2, i.e., $c_{12} \approx (w_1/w_2)$. This can be generalized and the statements below can be concluded:

- $c_{ij} \approx (w_i/w_j), i, j = 1, 2, \ldots, n$

- $c_{ii} \approx (w_i/w_i) = 1,\ i = 1, 2, \ldots, n$

- If $c_{ij} = \alpha,\ \alpha \neq 0$, then $c_{ji} = 1/\alpha,\ i = 1,\ 2,\ \ldots,\ n$
- If C_i is more important than C_j, then $c_{ij} \cong (w_i/w_j) > 1$

Table 1 Fuzzy Scale of valuation in the pair-wise comparison process [25]

Labels	Verbal judgments of preferences between criterion i and criterion j	Triangular fuzzy scale and reciprocals
(II)	Ci and Cj are equally important	(1, 1, 1)/(1,1,1)
(M + I)	Ci is slightly more important than Cj	(2, 3, 4)/(1/4,1/3,1/2)
(+I)	Ci is strongly more important than Cj	(4, 5, 6)/(1/6,1/5,1/4)
(Mu + I)	Ci is very strongly more important than Cj	(6, 7, 8)/(1/8,1/7,1/6)
(Ex + I)	Ci is extremely more important than Cj	(8, 9, 9)/(1/9,1/9,1/8)

This implies that the matrix C should be a positive and reciprocal matrix with 1's on the main diagonal. Hence, the decision maker only needs to provide value judgments in the upper triangle of the matrix. The values assigned to c_{ij} according to the Saaty scale usually lie in the interval of 1-9 or their reciprocals.

It can be shown that the number of judgments (L) needed in the upper triangle of the matrix is:

$$L = n(n-1)/2 \tag{7}$$

where n is the size of the matrix C.

As the reader can observe, there are both qualitative and quantitative criteria, so it is necessary to transform the Saaty's scale to fuzzy numbers. Therefore, Table 1 presents the decision-maker's linguistic preferences in the fuzzy pairwise comparison process.

The vector of weights is the eigenvector corresponding to the maximum eigenvalue "λ_{max}" of the matrix C. The traditional eigenvector method of estimating weights in AHP yields a way of measuring the consistency of the referee's preferences arranged in the comparison matrix.

In AHP problems, where the values are fuzzy not crisp, instead of λ using the eigenvector as an estimator of the weight, we will use the geometric normalized average, expressed by the following expression:

$$w_i = \frac{\left(\prod_{j=1}^{n}\left(a_{ij}, b_{ij}, c_{ij}\right)\right)^{1/n}}{\sum_{i=1}^{m}\left(\prod_{j=1}^{n}\left(a_{ij}, b_{ij}, c_{ij}\right)\right)^{1/n}} \tag{8}$$

where, (a_{ij}, b_{ij}, c_{ij}) is a fuzzy number.

Additionally, to obtain the weight vector, the normalizing operation must be used; this will be achieved through expression (9).

$$\left(w_{c_{ia}}, w_{c_{ib}}, w_{c_{ic}}\right) = \left[\frac{c_{ia}}{\sum_{i=1}^{n} c_{ic}}, \frac{c_{ib}}{\sum_{i=1}^{n} c_{ib}}, \frac{c_{ic}}{\sum_{i=1}^{n} c_{ia}}\right] \tag{9}$$

4 Determining the Criteria Importance

Not all the criteria which have influence in this kind of decision problems have the same importance. Besides, although there are decision problems that could be similar, the selection of the criteria depend of the specific necessities of each country. Therefore, not only it is important to carry out an appropriate selection of criteria, but also to choose the way of obtaining their weights. For instance, previous studies [22, 26] have determined the weights of criteria via direct assignment. However, in this study, the way of obtaining these weights has been through preferences of an experts group.

The group of experts involved in the decision process consisted of six experts specialized in this specific field (three trained test pilots and three flight instructors of the Spanish Air Force).

According to expression (7), 66 questions should be answered by each one of the experts. Despite the huge amount of work needed, it is possible that some inconsistent matrices can be generated. In order to decrease the inconsistency in this specific case study and to reduce the amount of work required for each expert, we reduced the number of questions in such a way that no loss of relevant information is produced [27, 28]. Therefore, we propose an alternative method, which only requires making $(n - 1)$ comparisons. For that purpose, a questionnaire similar to [29] was carried out. This questionnaire also allow us to reduce the uncertainty and imprecision in the proposed problem.

4.1 Fuzzy-Delphi-AHP Survey

The methodology used for the extraction of the experts' knowledge is a pseudo-Delphi technique, since the members who are part of the decision-making do not interact at any time. In order to do this, a series of questionnaires were distributed among the six participants in this process so that they could choose the answers they considered most appropriate, in order to reduce the uncertainty and vagueness involved with the problem presented.

The questionnaire designed has two clearly different parts. The first one consists of the presentation of the decision problem where the variables employed and the work methods to carry out are detailed. The experts were asked if the approach made to solve the problem was suitable and if they agreed with it. The six experts gave an affirmative answer and therefore it was possible to carry on with the survey (second part of the questionnaire).

It is known that if one criterion is more important than another, it should be considered that said criterion has a greater weight than the other. Therefore, the rest of the survey was focused on the following group of questions:

Table 2 Order of importance of criteria for each of the experts

E_1	$C_{11} = C_{10} > C_9 > C_{12} = C_1 > C_2 = C_3 = C_7 = C_8 > C_4 = C_5 = C_6$
E_2	$C_9 = C_{10} = C_{11} > C_3 = C_7 = C_8 > C_4 = C_1 = C_2 = C_5 = C_6 > C_{12}$
E_3	$C_{11} > C_{10} = C_9 = C_{12} > C_4 > C_3 = C_5 = C_2 = C_7 > C_6 = C_1 = C_8$
E_4	$C_{11} > C_{12} = C_7 = C_9 > C_{10} = C_8 = C_2 > C_1 = C_3 = C_4 > C_5 = C_6$
E_5	$C_{11} > C_9 = C_{10} > C_1 = C_2 = C_3 = C_4 = C_5 = C_6 = C_7 = C_8 = C_{12}$
E_6	$C_{11} > C_9 = C_{10} = C_{12} = C_2 = C_3 > C_7 = C_8 = C_5 > C_6 = C_4 = C_1$

a. ***Do you think that the twelve criteria considered have the same importance?***

If the answer is affirmative, the weight associated with criterion C_j is $w_j = 1/m$, $j = 1,2,\ldots,m$. If on the contrary, experts consider that not all the criteria have the same importance, then it is appropriate to proceed to the next question of the survey.

The next step will be to find the extent to which one criterion is more important than another, this degree of importance will be analyzed to be able to assign a weight to each criterion. For example, when indicating that a particular criterion has a higher weight than the rest of the criteria, it is declared that this is the most important criterion. Forthwith, the weights of the criteria will be used to quantify their importance.

The six experts have considered that certain criteria should have a greater weight than others. Therefore, those weights need to be determined.

b. ***Write the order of importance among the twelve criteria*** (Table 2).

As can be seen in Table 2, the six experts believe the importance of the criteria to be different, although they differ in the order of importance of the criteria. Analyzing the above table, experts indicate that criteria C_9, C_{10} and C_{11} are the most important criteria. Due to that, these criteria will have larger weights.

Once the expert has indicated the order of importance, the next question would be considered:

c. ***Compare the criterion chosen in first place with respect to that considered secondly and successively, using the following labels, {(II), (M +), (+I), (Mu + I), (Ex + I)} according to the meanings in*** Table 1.

To determine the weights of the criteria, as has been discussed, a pair-wise comparison has been made. Using Expert 1 as an example, in Fig. 1 his appreciation by pair-wise comparison is shown.

The meaning is as follows: criterion C_{11} is extremely more important than C_4, C_5 and C_6, with respect to C_2, C_3, C_7 and C_8 it is very strongly more important,

$$
\begin{array}{cc}
 & \begin{array}{cccccccccccc} C_{11} & C_{10} & C_9 & C_{12} & C_1 & C_2 & C_3 & C_7 & C_8 & C_4 & C_5 & C_6 \end{array} \\
C_{11} & \left[\begin{array}{cccccccccccc} II & II & M+I & +I & +I & Mu+I & Mu+I & Mu+I & Mu+I & Ex+I & Ex+I & Ex+I \end{array}\right]
\end{array}
$$

Fig. 1 Valuations given by E_1

$$\begin{array}{c} \\ C_{11} \end{array} \begin{array}{cccccccccccc} C_{11} & C_{10} & C_{9} & C_{12} & C_{1} & C_{2} & C_{3} & C_{7} & C_{8} & C_{4} & C_{5} & C_{6} \\ \end{array}$$

$$C_{11}\ \begin{bmatrix} (1,1,1) & (1,1,1) & (2,3,4) & (4,5,6) & (4,5,6) & (6,7,8) & (6,7,8) & (6,7,8) & (6,7,8) & (8,9,9) & (8,9,9) & (8,9,9) \end{bmatrix}$$

Fig. 2 Matrix of decision making for E_1

$$\begin{array}{c} \\ C_{11} \\ C_{10} \\ C_{9} \\ C_{12} \\ C_{1} \\ C_{2} \\ C_{3} \\ C_{7} \\ C_{8} \\ C_{4} \\ C_{5} \\ C_{6} \end{array} \begin{array}{c} C_{11} \\ \left[\begin{array}{c} (1,1,1) \\ (1,1,1) \\ (1/4,1/3,1/2) \\ (1/6,1/5,1/4) \\ (1/6,1/5,1/4) \\ (1/8,1/7,1/6) \\ (1/8,1/7,1/6) \\ (1/8,1/7,1/6) \\ (1/8,1/7,1/6) \\ (1/9,1/9,1/8) \\ (1/9,1/9,1/8) \\ (1/9,1/9,1/8) \end{array}\right] \end{array} = \begin{bmatrix} (0.247, 0.275, 0.293) \\ (0.247, 0.275, 0.293) \\ (0.062, 0.092, 0.146) \\ (0.041, 0.055, 0.073) \\ (0.041, 0.055, 0.073) \\ (0.031, 0.039, 0.049) \\ (0.031, 0.039, 0.049) \\ (0.031, 0.039, 0.049) \\ (0.031, 0.039, 0.049) \\ (0.027, 0.031, 0.037) \\ (0.027, 0.031, 0.037) \\ (0.027, 0.031, 0.037) \end{bmatrix}$$

Fig. 3 Criteria weight for E_1

with respect to C_{12} and C_1 it is strongly more important, with respect to C_9 it is slightly more important and with respect to C_{10} is equally important.

This, translated to the fuzzy numbers according to Table 1, gives the results shown in Fig. 2.

Taking into account [30] and operation (9), the weights of the considered criteria are obtained (Fig. 3).

The information detailed above for E_1 would also be carried out for the other experts. The normalized weights associated with the corresponding criterion C_j, $j = 1,2,\ldots,12$ given by each of the experts can be seen in Table 3.

Analyzing the above table, criterion C_{11} (security systems) has the maximum score for each of the experts; this criterion is equally important to criterion C_{10} (flying and handling qualities) for expert 1, and equally important to criteria C_{10} and C_9 (human factors) for expert 2. This expert also considers as the second most important criteria C_3 (stalling speed), C_7 (take-off distance) and C_8 (landing distance). Conversely, the least important criterion for this expert is C_{12} (tactical capability).

The weights of the criteria for expert 3 are similar. According to this expert, criterion C_{11} is also the highest rated, with the second most important criteria being C_9, C_{10} and C_{12}. The least important criteria are C_1 (service ceiling), C_6 (negative limit load factor) and C_8.

Table 3 Weights of criteria for the six experts (heterogeneous aggregations)

	Normalized (Expert 1)			Normalized (Expert 2)			Normalized (Expert 3)		
C1	[0.041	0.055	0.073]	[0.023	0.030	0.037]	[0.029	0.036	0.045]
C2	[0.031	0.039	0.049]	[0.023	0.030	0.037]	[0.033	0.046	0.061]
C3	[0.031	0.039	0.049]	[0.046	0.069	0.111]	[0.033	0.046	0.061]
C4	[0.027	0.031	0.037]	[0.023	0.030	0.037]	[0.044	0.064	0.091]
C5	[0.027	0.031	0.037]	[0.023	0.030	0.037]	[0.033	0.046	0.061]
C6	[0.027	0.031	0.037]	[0.023	0.030	0.037]	[0.029	0.036	0.045]
C7	[0.031	0.039	0.049]	[0.046	0.069	0.111]	[0.033	0.046	0.061]
C8	[0.031	0.039	0.049]	[0.046	0.069	0.111]	[0.029	0.036	0.045]
C9	[0.062	0.092	0.146]	[0.183	0.207	0.223]	[0.066	0.107	0.182]
C10	[0.247	0.275	0.293]	[0.183	0.207	0.223]	[0.066	0.107	0.182]
C11	[0.247	0.275	0.293]	[0.183	0.207	0.223]	[0.264	0.322	0.364]
C12	[0.041	0.055	0.073]	[0.020	0.023	0.028]	[0.066	0.107	0.182]
	Normalized (Expert 4)			Normalized (Expert 5)			Normalized (Expert 6)		
C1	[0.031	0.044	0.059]	[0.039	0.058	0.083]	[0.036	0.040	0.049]
C2	[0.042	0.062	0.088]	[0.039	0.058	0.083]	[0.053	0.072	0.098]
C3	[0.031	0.044	0.059]	[0.039	0.058	0.083]	[0.053	0.072	0.098]
C4	[0.031	0.044	0.059]	[0.039	0.058	0.083]	[0.036	0.040	0.049]
C5	[0.028	0.034	0.044]	[0.039	0.058	0.083]	[0.040	0.052	0.066]
C6	[0.028	0.034	0.044]	[0.039	0.058	0.083]	[0.036	0.040	0.049]
C7	[0.063	0.103	0.176]	[0.039	0.058	0.083]	[0.040	0.052	0.066]
C8	[0.042	0.062	0.088]	[0.039	0.058	0.083]	[0.040	0.052	0.066]
C9	[0.063	0.103	0.176]	[0.059	0.096	0.167]	[0.053	0.072	0.098]
C10	[0.042	0.062	0.088]	[0.059	0.096	0.167]	[0.053	0.072	0.098]
C11	[0.250	0.308	0.351]	[0.235	0.288	0.333]	[0.320	0.362	0.393]
C12	[0.063	0.103	0.176]	[0.039	0.058	0.083]	[0.053	0.072	0.098]

Apart from criteria C_9 and C_{12}, expert 4 also considers C_7 as the second most important criterion. The least important criteria for this expert are C_5 and C_6 (positive and negative limit load factors).

According to expert 5, criteria C_9 and C_{10} are the second most important criteria, while for that expert the remaining criteria have the same importance.

Expert 6 estimates that there is a criteria group consisting of C_2 (cruising speed), C_3, C_9, C_{10} and C_{12} which have the following score after that of the highest criterion (C_{11}). The least important criteria are C_1, C_4 (endurance) and C_6.

In order to unify the weights of the obtained criteria and to stablish a specific weight for each one of the criteria, a homogeneous aggregation will be carried out. i.e., all experts are equally important in the decision, as a measure of aggregation the arithmetic average will be used (expression 10).

Table 4 Weights of criteria through experts' homogeneous aggregation

Experts' homogeneous aggregation			
C_1	[0.0332	0.0437	0.0578]
C_2	[0.0368	0.0511	0.0693]
C_3	[0.0389	0.0547	0.0768]
C_4	[0.0334	0.0444	0.0593]
C_5	[0.0317	0.0416	0.0545]
C_6	[0.0304	0.0380	0.0493]
C_7	[0.0419	0.0611	0.0909]
C_8	[0.0378	0.0525	0.0737]
C_9	[0.0809	0.1129	0.1653]
C_{10}	[0.1084	0.1366	0.1750]
C_{11}	[0.2499	0.2937	0.3262]
C_{12}	[0.0471	0.0697	0.1067]

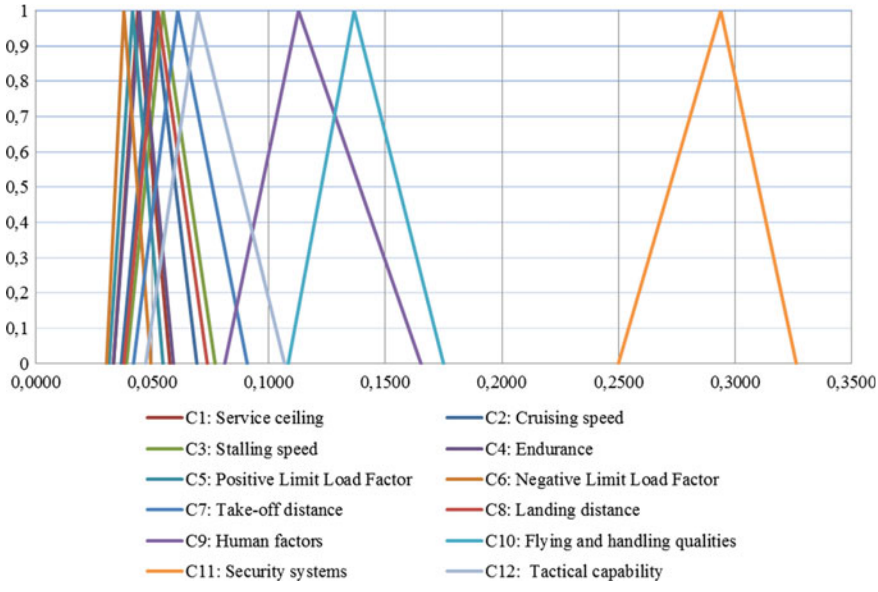


Fig. 4 Graphic representation (experts' homogeneous aggregation)

$$(\bar{X}_{ia}, \bar{X}_{ib}, \bar{X}_{ic}) = \left[\frac{\sum_{i=1}^{n} X_{ia}}{n}, \frac{\sum_{i=1}^{n} X_{ib}}{n}, \frac{\sum_{i=1}^{n} X_{ic}}{n}\right] \quad (10)$$

By the homogeneous aggregations indicated, the weights of the criteria will be obtained, taking into account the entire decision-making group. Therefore, the values obtained for the selection problem of the best military training aircraft are those indicated in Table 4 and Fig. 4.

Through homogeneous aggregation it is observed that the most important criteria are C_{11} (security systems), C_{10} (flying and handling qualities) and C_9 (human factors). According to experts 1, 2, 3, 5 and 6 these criteria are also the most important criteria. The only expert who lightly differs of the rest of experts is expert 4. This expert indicates as the second criteria in importance order the criteria C_7 (take-off distance), C_9 and C_{12} (tactical capability) while criterion C_{10} is moved to the third position.

The following criteria group in importance is comprised of two criteria; C_7 (take-off distance), and C_{12} (tactical capability) which are the criteria that expert 4 located in the second position. Whereas the least important criteria are C_5 and C_6 (positive and negative limit load factors).

5 Conclusions

With respect to the applied methodology, it is worth highlighting that, carrying out the extraction of knowledge from an experts group in this specific field (trained test pilots and flight instructors of the Spanish Air Force) has allowed to combine the Delphi method and the fuzzy logic techniques with a well-known decision making tool like AHP methodology.

Furthermore, it has not only been possible to select and define a list of criteria which influence the selection problem, but also to obtain their coefficients of importance through the AHP methodology.

Through the homogeneous aggregation, it is observed that the most important criteria when selecting the best military training aircraft are C_{11} (security systems), C_{10} (flying and handling qualities) and C_9 (human factors).

Finally, it should be emphasized that the aforementioned criteria constitute the group of relevant criteria which should be taken into account in order to preserve the security or decrease of risk during the training, to extend this work, a further study regarding additional relevant criteria, such as economic aspects or even institutional factors, should be carried out.

Acknowledgements The authors acknowledge support through the projects TIN2014-55024-P from the Spanish Ministry of Economy and Competitiveness, and P11-TIC-8001 from the Consejería de Economía, Innovación y Ciencia of Junta de Andalucía (both including FEDER funds from the EU).

References

1. Oroumieh, M.A.A., Malaek, S.M.B., Ashrafizaadeh, M., Taheri, S.M.: Aircraft design cycle time reduction using artificial intelligence. Aerosp. Sci. Technol. **26**, 244–258 (2013)

2. Wei, S.D., Xing, G.P. Sun, D.X., Gao, K., Liu, Y.W.: Research on SPA-based approaches and application of the evaluation for maintenance quality of military aircraft. 978-1-4577-1232-6/11. IEEE. New York (2011)
3. Cheng, C.H., Yang, K.L., Hwang, C.L.: Evaluating attack helicopters by AHP based on linguistic variable weight. Eur. J. Oper. Res. **116**, 423–435 (1999)
4. Israeli, A.A., Mehrez, A., Bochen, D., Hakim, D.: Justification of global positioning systems purchase using the analytic hierarchical process—The case of the Israeli Defense Force. Technovation **18**, 409–424 (1998)
5. Crary, M., Nozick, L.K., Whitaker, L.R.: Sizing the US destroyer fleet. Eur. J. Oper. Res. 136: 680–695 (2002)
6. Ullah, R., Zhou, D.Q., Zhou, P., Hussain, M., Sohail, M.A.: An approach for space launch vehicle conceptual design and multi-attribute evaluation. Aerosp. Sci. Technol. **25**, 65–74 (2013)
7. Sánchez-Lozano, J.M., Teruel-Solano, J., Soto-Elvira, P.L., García-Cascales, M.S.: Geographical information systems (GIS) and multi-criteria decision making (MCDM) methods for the evaluation of solar farms locations: case study in south-eastern Spain. Renew. Sustain. Energy Rev. **24**, 544–556 (2013)
8. Gómez-López, M.D., Bayo, J., García-Cascales, M.S., Angosto, J.M.: Decision support in disinfection technologies for treated wastewater reuse. J. Clean. Prod. **17**, 1504–1511 (2009)
9. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
10. Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, New Jersey (1995)
11. Hajela, P.: Soft computing in multidisciplinary aerospace design-new directions for research. Prog. Aerosp. Sci. **38**, 1–21 (2002)
12. Cheng, C.H.: Evaluating weapon systems using ranking fuzzy numbers. Fuzzy Sets Syst. **107**, 25–35 (1999)
13. Hossain, A., Rahman, A., Hossen, J., Iqbal, A.K.M.P., Zahirul, M.I.: Prediction of aerodynamic characteristics of an aircraft model with and without winglet using fuzzy logic technique. Aerosp. Sci. Technol. **15**, 595–605 (2011)
14. Saaty, T.L.: The Analytic Hierarchy Process. McGraw Hill, New-York (1980)
15. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Methods and Applications. Springer, Berlin (1981)
16. Roy, B.: Classement et choix en presence de points de vue multiples (la method ELECTRE). Revue informatique et recherché opérationnelle **8**, 57–75 (1968)
17. Winchester, J.: Modern Military Aircraft (Aviation Factfile. The). Amber Books Ltd., London (2004)
18. Aeronaves operativas del ejército del aire español. Ministerio de defensa. http://www.ejercitodelaire.mde.es/ea/. Accessed 3 June 2014
19. Noor, A.K., Venneri, S.L.: Future Aeronautical and Space Systems. Progress in Astronautics and Aeronautics. 72. American Institute of Aeronautics and Astronautics. Virginia (1997)
20. Eshelby, M.E.: Aircraft performance—Theory and practice. Elsevier, London (2000)
21. Newberry, C.F.: The conceptual design of deck-launched waverider-configured aircraft. Aircr. Des. **1**, 159–191 (1998)
22. Wang, T.C., Chang, T.H.: Application of TOPSIS in evaluating initial training aircraft under a fuzzy environment. Expert Syst. Appl. **33**, 870–880 (2007)
23. Bellman, R.E., Zadeh, L.A.: Decision-Making in a fuzzy environment. Washington. D.C. National Aeronautics and Space Administration (NASA CR-1594) (1970)
24. García-Cascales, M.S., Lamata, M.T.: Multi-criteria analysis for a maintenance management problem in an engine factory: rational choice. J. Intell. Manuf. **22**, 779–788 (2011)
25. Saaty, T.L.: Group decision making and the AHP. Springer, New York (1989)

26. Wang, J., Fan, K., Su, Y., Liang, S., Wang, W.: Air combat effectiveness assessment of military aircraft using a fuzzy AHP and TOPSIS methodology. Asia Simulation Conference—7th International Conference on System Simulation and Scientific Computing (2008)
27. Sánchez-Lozano, J.M., García-Cascales, M.S., Lamata, M.T., Sierra, C.: Decision Criteria for Optimal Location of Wind Farms, Exploring Innovative and Successful Applications of Soft Computing. IGI Global Ed (2013)
28. Sánchez-Lozano, J.M., García-Cascales, M.S., Lamata, M.T.: Identification and selection of potential sites for onshore wind farms development in Region of Murcia, Spain. Energy 2014; **73**, 311–324 (2014)
29. Garcia-Cascales, M.S., Lamata, M.T., Sanchez-Lozano, J.M.: Evaluation of photovoltaic cells in a multi-criteria decision making process. Ann. Oper. Res. **199**, 373–391 (2012)
30. Zadeh, L.A., Kacprycz, J.: Computing with Words in Information/Intelligent Systems. Part 1. Physica-Verlag (Springer-Verlag), Heidelberg and New York (1999)

Participatory Search in Evolutionary Fuzzy Modeling

Yi Ling Liu and Fernando Gomide

Abstract Search is one of the most useful procedures employed in numerous situations such as optimization, machine learning, information processing and retrieval. This chapter introduces participatory search, a class of population-based search algorithms constructed upon the participatory learning paradigm. Participatory search relies on search mechanisms that progress forming pools of compatible individuals. The individual that is the most compatible with the best individual is always kept in the current population. Random immigrants are added to complete the population at each algorithm step. Different types of recombination are possible. The first is a convex combination, arithmetic-like recombination modulated by the compatibility between individuals. The second is a recombination mechanism based on selective transfer. Mutation is an instance of differential variation modulated by compatibility between selected and recombined individuals. Applications concerning development of fuzzy rule-based models from actual data illustrate the potential of the algorithms. The performance of the models produced by participatory search algorithms are compared with a state of the art genetic fuzzy system. Experimental results show that the participatory search algorithm with arithmetic-like recombination performs better than the remaining ones.

1 Introduction

The interest in evolutionary procedures to develop fuzzy systems from data has gained considerable attention in the last decade. Evolutionary fuzzy systems are fuzzy systems with added evolutionary components. An important instance of evolutionary fuzzy systems is genetic fuzzy systems (GFS). GFS combine fuzzy systems

Y.L. Liu (✉) · F. Gomide
School of Electrical and Computer Engineering, University of Campinas,
Sao Paulo, Brazil
e-mail: yiling155@gmail.com

F. Gomide
e-mail: gomide@dca.fee.unicamp.br

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_11

with genetic algorithms [1] to solve complex classification, approximation, nonlinear modeling and control problems.

As it is well known, genetic algorithm (GA) is a population-based stochastic search procedure whose idea is to evolve a population of individuals using selection, recombination, and mutation operations working in sequence during several steps called generations [2]. A fitness function distinguishes the ability of an individual to remain in the next population. The better the value of the fitness function achieved by an individual, the higher is its chance to survive. This is the survival of the fittest saga. Individuals, candidate solutions of a problem, are points in the search space. Differently from GA, differential evolution [3] creates new candidate solutions combining the existing ones via mutation, recombination, and selection working in sequence during several generations. DE keeps whichever candidate solution that achieves the highest performance.

In [4] we read the following: *In actual survival of the fittest saga, there appears to be additional processes going on. In particular, the objective function in addition to be determined by some external requirement is often affected by the population itself.*

An approach that has been devised mimic the effect that a population itself has in its evolution is participatory learning [5]. The key idea of participatory learning is to account for compatibility between observations and current state of the learner. As it will be shown late, selection and variation operators such as recombination and mutation can designed to account for the compatibility between the individuals of a population. Compatibility and similarity have been shown to be effective in evolutionary computation [6–9].

This chapter addresses a new class of population-based search algorithms based on participatory learning. In common with other types of evolutionary algorithms, participatory search operates with a population of solutions, rather than with a single solution at a step, and employs procedures to combine these solutions to create new ones. Participatory search algorithms are novel instances of evolutionary algorithms because they do not need to assume that evolutionary approaches must necessarily be based on randomization [10, 11] though they are compatible with randomized implementations. Participatory search algorithms embody principles that are still not used by other evolutionary approaches, and that prove advantageous to solve a variety of complex optimization and design tasks.

The performance of the participatory algorithms is evaluated using actual data and compared with a state of the art genetic fuzzy system approach developed in [1]. Computational results show that the participatory search algorithm with arithmetical-like recombination performs better than the GFS approach.

After this introduction the chapter proceeds as follows. Section 2 briefly reviews genetic fuzzy systems. Section 3 reminds the concept of participatory learning. Section 4 introduces the participatory search operators: selection, selective transfer, arithmetic-like recombination and mutation operators. The search algorithms summarized in Sect. 5. Section 6 evaluates the performance of the participatory search algorithms against state of the art genetic fuzzy systems approaches. Section 7 concludes the chapter and list issues that deserve further development.

2 Genetic Fuzzy Systems

This section gives a brief overview of genetic fuzzy systems (GFS) and their applications. The focus is on genetic fuzzy rule-based systems (GFRBS), one of the most important types of GFS. The structure of GFRBS is summarized in Fig. 1.

GFRBS is a fuzzy rule-based system enhanced by genetic algorithms. A fuzzy rule-based system (FRBS) is composed by a knowledge base (KB) that encodes the knowledge of a target model. The KB contains two main components, a data base and a fuzzy rule base. The data base (DB) stores data that characterize the linguistic variables used by the fuzzy rules, the membership functions that define the semantics of the linguistic labels, and the parameters of the model. The fuzzy rule base (RB) is a collection of fuzzy if-then rules. Other three components complete fuzzy rule-based models. The first is a fuzzification module to serve as an input interface with the fuzzy reasoning process. The second is an inference engine that performs fuzzy reasoning. The third is a defuzzification output interface module to convert a fuzzy output into a representative pointwise output. An effective approach to construct the KB of an FRBS is to simultaneously develop the DB and the RB within the same process, but in two steps such as in embedded GFRBS learning. Embedded GFRBS is a scheme to learn the DB using simultaneously a simple method to derive a RB for each DB.

Embedded GFRBS does not necessarily provide simple, transparent, and competitive models in terms of the generalization capability. They may not scale well in terms of processing time and memory, two essential requirements especially in high-dimensional, large-scale, and complex problem solving. These issues are addressed in [1] where a way to reduce the search space in an embedded genetic DB learning framework is suggested. Lateral displacement of fuzzy partitions using a unique parameter for all membership functions of each linguistic variable is one of the mechanisms adopted to reduce search space complexity. The idea is to prescreen promising partitions to avoid overfitting and to maintain coverage and semantic soundness

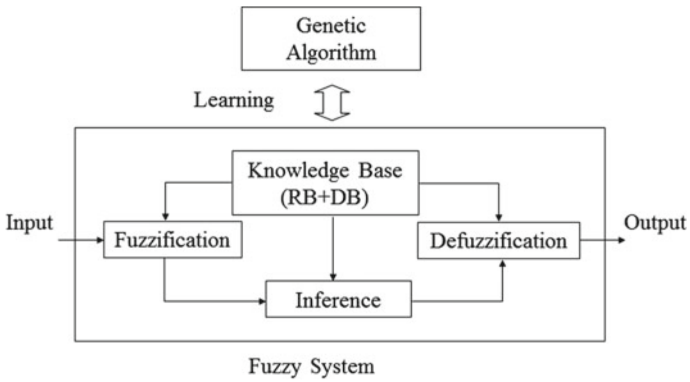


Fig. 1 Genetic fuzzy rule-based system

of the fuzzy partitions. The evolutionary algorithm also includes incest prevention, restarting, and rule-cropping in the RB generation process to improve convergence. Despite the use of mechanisms to manage dimensionality, the algorithm does not scale up on the number of data in datasets. A way to deal with scalability is to avoid large percentage of samples, and to estimate errors using a reduced subset. A post-processing step may further refine the algorithm.

Application examples of GFS are many. For example, [12] addresses a multi-objective optimization in which a fuzzy controller regulates the selection procedure and fitness function of genetic algorithms. Optimization is used to develop timetables of railway networks aiming at reducing passenger waiting time when switching trains, while at the same time, minimizing the cost of new investments to improve the necessary infrastructure. The result of the genetic optimization is a cost-benefit curve that shows the effect of investments on the accumulated passenger waiting time and trade-offs between both criteria. In [13] the aim is to optimize trip time and energy consumption of a high-speed railway with fuzzy c-means clustering and genetic algorithm. The method is used to develop a control strategy for a high-speed train line. An economical train runs with a trip time margin of less than 7% and an energy saving of 5% is reported. A model to relate the total length of low voltage line installed in a rural town with the number of people in the town and the mean of the distances from the center of the town to three furthest clients is discussed in [14]. The authors compare the training and test set error achieved by different modeling techniques for low line value estimation.

3 Participatory Learning

Participatory learning appeared in [5] as a process of learning that depends on what has already been learned. A central issue in the idea of participatory learning is that data have the greatest impact in causing learning or knowledge revision when they are compatible with the current knowledge. Learning occurs in an environment in which the current knowledge participates in the process of learning about itself. Clearly, a fundamental factor of participatory learning is the compatibility degree between input data and current knowledge. The current knowledge, denoted by $v(t)$, in addition to provide a standard against which input data $z(t)$ is compared with, directly affects the learning process. This is the participatory nature of learning process. High compatibility between the current knowledge and current input data opens the system for learning. In PL, this enhancement is expressed by the compatibility degree. A facility is provided to measure the confidence in the current knowledge structure. If a long sequence of input data have low compatibility with current knowledge, it may be the case that what has been learned so far is mistaken, not the data. This is seen as a form of stimulation called arousal. Participatory learning includes an arousal mechanism to monitor the performance of the learning process by watching at the values of the compatibility degrees of the current knowledge with inputs. Monitoring information is fed back in terms of an arousal index that subsequently affects the learning process.

The instance of participatory learning we explore in this chapter uses the compatibility degree between current knowledge and current input data to update knowledge employing the following procedure [5, 15]:

$$v(t+1) = v(t) + \alpha \rho_t (z(t) - v(t)) \tag{1}$$

where $v(t)$ and $z(t)$ are n-dimensional vectors that denote the current knowledge and current input data, respectively. Assume, without loss of generality, that $v(t), z(t) \in [0,1]^n$. The parameter $\alpha \in [0,1]$ is the basic learning rate and $\rho_t \in [0,1]$ is the compatibility degree between $v(t)$ and $z(t)$ at step t. The product of the basic learning rate by the compatibility degree produces the actual learning rate. If an input is far from the current knowledge, then the value of the corresponding compatibility degree is small and the input is filtered. The actual learning rate is lowered by the compatibility degree. This means that if input data are too conflicting with the current knowledge, then they are discounted [5]. Lower values of actual learning rates avoid fluctuations due to values of input data which do not agree with current knowledge. As it will be shown shortly, (1) induces one of the recombination operators of participatory search algorithms.

The mechanism to monitor compatibility degrees during learning is the arousal index. The arousal index enters in the basic PL update formula (1) as follows

$$v(t+1) = v(t) + \alpha \rho_t^{1-a_t} (z(t) - v(t)) \tag{2}$$

where $a_t \in [0,1]$ is the arousal index at t.

One way to compute the compatibility degree ρ at step t is

$$\rho_t = 1 - \frac{1}{n} \sum_{k=1}^{n} |z_k(t) - v_k(t)|. \tag{3}$$

In (3) ρ_t is the complement of the average absolute difference between input information $z(t)$ and current knowledge $v(t)$. In a more general sense, ρ_t may be seen to be a measure of similarity between $z(t)$ and $v(t)$. If $\rho_t = 0$, then $v(t+1) = v(t)$ and the current input $z(t)$ is completely incompatible with the current knowledge $v(t)$. This condition means that the system is not open to any learning from the current information. On the other hand, if $\rho_t = 1$, then $v(t+1) = z(t)$. In this case input information is in complete agreement with the current knowledge and the system is fully open to learn.

Arousal can be seen as the complement of the confidence in the current knowledge. A simple procedure is to update the arousal index a at step t is

$$a_{t+1} = (1-\beta) a_t + \beta (1 - \rho_{t+1}) \tag{4}$$

where $\beta \in [0, 1]$ controls the rate of change of arousal. The higher a_t, the less confident is the learning system in current knowledge. If $\rho_{t+1} = 1$, then we have a highly compatible input and the arousal index decreases. On the other hand, if $\rho_{t+1} = 0$, then input information compatibility is low and the arousal index increases.

The notion of compatibility degree enters in participatory search algorithms during the formation of pools of individuals for selection, recombination, and mutation. The pools are assembled from two populations S^t and $S^{t'}$. The individuals of $S^{t'}$ are those of S^t which are the most compatibles, one to one. Selection uses compatibility to choose those individuals from the pool that are closer to current best individual. Recombination is done pairwise between individuals of the mating pool, modulated by their compatibility degrees and arousal indexes. Mutation adds a variation to the current best individual proportional to the difference between the selected and recombined individuals modulated by the corresponding compatibility degrees. The effect of compatibility is to encourage selection and recombination of similar mates from which good offspring are likely to be produced, as indicated in [9].

4 Participatory Search Operators

The main construct elements of search algorithms are the representation, search operators, fitness function, and initial solution. These elements are relevant for all types of population-based algorithms. The remaining element is the search strategy. Representation concerns encoding mechanisms that maps problems solutions to strings. Representations allow definitions of search operators and of the search space. The search strategy defines types of intensification and diversification mechanisms.

In what follows we assume that a populations is a finite set of strings.

4.1 Selection

Let S be a set of N strings of fixed length n, and $s, s^{'} \in S$ be two individuals, $s^{'}$ distinct of s, such that

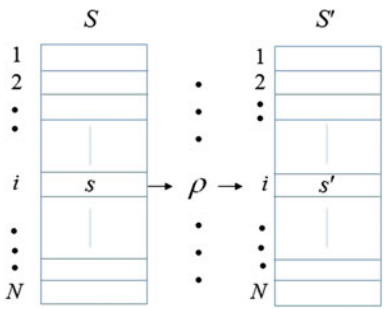
$$s' = argmax_{r \in S}(\rho(s, r)) \tag{5}$$

where

$$\rho(s, r) = 1 - \frac{1}{n} \sum_{k=1}^{n} |s_k - r_k|, \tag{6}$$

and $s = (s_1, s_2, .., s_N)$ and $r = (r_1, r_2, .., r_N)$. Expression (5) means that $s^{'}$ is the individual of S whose compatibility degree with s is the largest. This procedure is repeated in sequence for each individual s of S to assemble a corresponding pool

Fig. 2 A population and its pool of N individuals



S' with N individuals. Notice that construction of the pool is biased by the compatibility degrees between the individuals of S. Figure 2 illustrates how the populations S and S' are assembled.

In participatory search algorithms, selection is done by computing the compatibility degrees between $s \in S$ and the corresponding $s' \in S'$ with the current best individual $best = s^*$, and picking the one that is the most compatible to assemble a population L of selected individuals, that is, the ones that are the closest to the current best individual. Formally,

$$s^* = argmin_{s \in S} f(s), \tag{7}$$

where f is the objective function.

More specifically, selection computes the compatibility degrees $\rho^s(s, s^*)$ and $\rho^{s'}(s', s^*)$ using

$$\rho^s = 1 - \frac{1}{n} \sum_{k=1}^{n} |s_k - s_k^*| \tag{8}$$

and

$$\rho^{s'} = 1 - \frac{1}{n} \sum_{k=1}^{n} |s'_k - s_k^*|, \tag{9}$$

and the individual whose compatibility degree is the largest, denoted by $p_{selected}$, is selected. That is, participatory selection proceeds according to the following rule

$$\text{if} \quad \rho^s \geq \rho^{s'} \quad \text{then} \quad p_{selected} = s \quad \text{else} \quad p_{selected} = s'. \tag{10}$$

Fig. 3 illustrates the process of selection.

Selection depends on the objective function $f(s)$, which identifies current best s^*, and on $\rho^s(s, s^*)$ and $\rho^{s'}(s', s^*)$ which measure the compatibility between s^* and the corresponding pair of individuals s and s' of the current pool. Jointly, f, ρ^s and $\rho^{s'}$ decide if an individual will be selected or not.

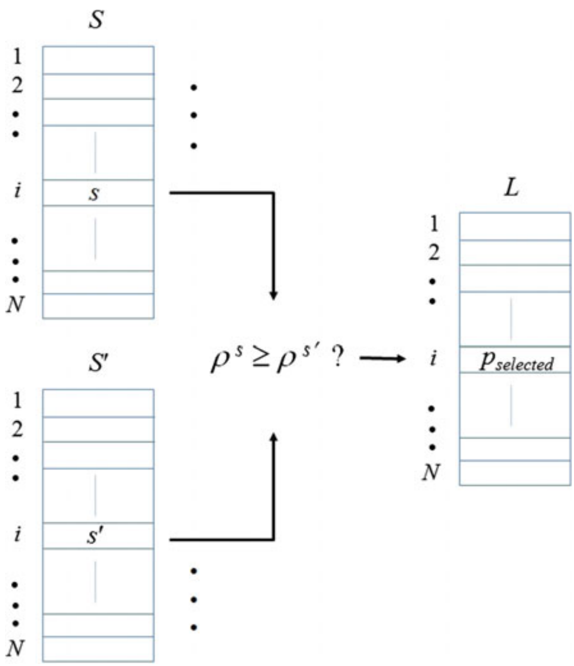


Fig. 3 Selection

4.2 Selective Transfer

During the last few years, we have witnessed a growing interest to use economic principles and models of learning in genetic algorithms. For instance, evolutionary processes have been used to model the adaptive behavior of a population of economic agents [16]. Here agents develop models of fitness to their environment in conjunction with the corresponding economic activities. Economists believe that behavior acquired through individual experience can be transmitted to future generations, and that learning changes the way to search the space in which evolution operates. This is an argument in favor of the interaction between the processes of evolution and learning. Since technical knowledge is distributed across the economic population, technological change can be viewed as a process of distributed learning. Here, the term learning is used in a broad sense, that is, there is no distinction between learning as propagation of knowledge through the population and the process of innovation, creation, and discovery. The distributed learning perspective helps to understand technological change and focus on the population suggests that an evolutionary perspective may be appropriate.

Birchenhall and Lin [16] claim that our knowledge and technology are modular, i.e., they can be decomposed into several components or modules. From the evolutionary computation point of view, they suggest that the crossover operator of genetic algorithms could be seen as a representative of modular imitation. To bring

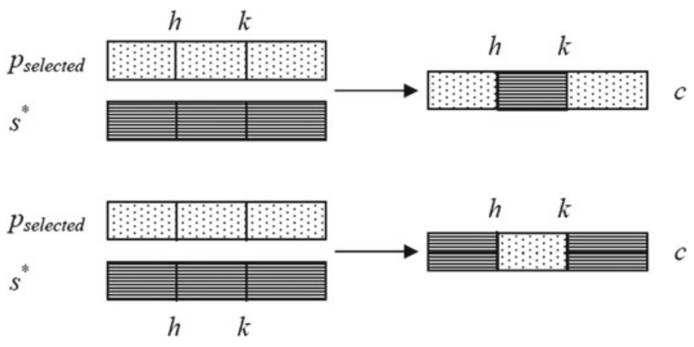


Fig. 4 Selective transfer

these ideas together, they advocate an algorithm that replaces selection and crossover operators by an operator based on selective transfer. Essentially, selective transfer is a filtered replacement of substrings from one string to another, without excluding the possibility that the entire sequence is copied [17]. Clearly, the selective transfer is similar to Holland crossover, but it is one-way transfer of strings, not on exchange of strings. The behavior selective transfer is likely to be very different from the combination of selection and crossover.

Assume that an individual $p_{selected}$ is selected using the objective function and compatibility. Two positions $h \leq k$ in the $p_{selected}$ string are chosen randomly, and a fair coin is tossed. If the coin turns head, then the substrings from $p_{selected}(h)$ to $p_{selected}(k)$ of $p_{selected}$ is replaced by the corresponding substrings from $s^*(h)$ to $s^*(k)$ of s^*. If the coin turns up tail, then the substrings from $p_{selected}(1)$ to $p_{selected}(h-1)$ and from $p_{selected}(k+1)$ to $p_{selected}(n)$ are replaced by the corresponding substrings of s^*. These steps are repeated for all individuals of L. Figure 4 illustrates the idea of selective transfer.

Despite similarity with crossover of the standard genetic algorithms, there are some differences. The most important one is that selective transfer uses one-way relocation of substrings, from the best individual to the one selected, and hence it is not a crossover. This is important because selective transfer is much more schemata destructive than the standard crossover [17].

4.3 Arithmetic Recombination

Arithmetic recombination emerges from the participatory learning update formula (2). To see this, notice that (2) can be rewritten as

$$\begin{aligned} v(t+1) &= v(t) + \alpha\rho_t^{(1-a_t)}(z(t) - v(t)) \\ &= (1 - \alpha\rho_t^{(1-a_t)})v(t) + \alpha\rho_t^{(1-a_t)}z(t). \end{aligned} \tag{11}$$

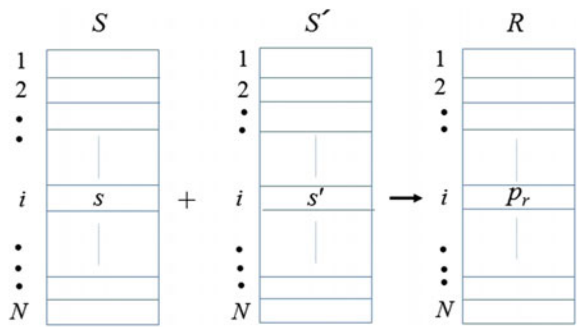


Fig. 5 Recombination

Let $\gamma = \alpha\rho_t^{(1-a_t)}$. Thus (11) becomes

$$v(t+1) = (1-\gamma)v(t) + \gamma z(t). \tag{12}$$

Expression (12) is of the following type

$$s_v(t+1) = (1-\delta)s_v(t) + \delta s_z(t) \tag{13}$$

where $\delta \in [0, 1]$. Notice that (13) is a convex combination of $s_v(t)$ and $s_z(t)$ whose result is the offspring $s_v(t+1)$. Interestingly (12) is similar to (13) and hence (12) is an arithmetic-like recombination. While parameter δ of (13) is either a constant or variable, depending on the age of population, the value γ of (12) is variable and modulated by compatibility and arousal.

Participatory recombination proceeds as in (12) to produce offspring p_r from individuals s and s' of pools S and S', respectively, as follows

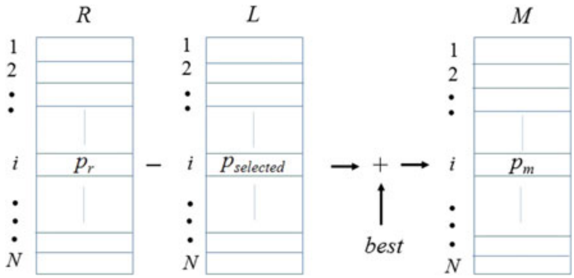
$$p_r = (1 - \alpha\rho_r^{(1-a)})s + \alpha\rho_r^{(1-a)}s'. \tag{14}$$

Figure 5 illustrates the process of participatory recombination. Sums in the figure are done on an individual basis, and should be understood from the point of view of the operation (14).

4.4 Mutation

There are many ways to do mutation in search algorithms. For example, consider a population of N individuals represented by n-dimensional vectors denoted by $s_{r_i,t}$ at generation t. Differential evolution, for instance, produces new individuals by adding the weighted differences between distinct vectors to a third vector [3]. For each vector $s_{r_i,t}$, $i = 1, 2, ..., N$, a mutated vector is generated using

Fig. 6 Mutation



$$s_{i,t+1} = s_{r_1,t} + \phi \cdot (s_{r_2,t} - s_{r_3,t}) \tag{15}$$

where $r_1, r_2, r_3 \in \{1, 2, ..., N\}$ are random indexes, and $\phi > 0$ is a parameter which controls the amount of the differential variation $(s_{r_2,t} - s_{r_3,t})$.

Mutation in participatory search is similar to differential evolution mutation. It produces a mutated individual p_m as follows

$$p_m = best + \rho_m^{1-a}(p_{selected} - p_r). \tag{16}$$

Fig. 6 illustrates the process of mutation.

In participatory mutation, the amount of the variation of the best individual $best = s^*$ is controlled by compatibility between the selected and recombined individuals, and the arousal index.

5 Participatory Search Algorithms

Let S^t be the set of N with strings of length n at step t. The participatory search algorithms (PSA) start with a population S^t at $t = 0$ with N randomly chosen individuals and, for each individual of S^t, the most compatible individual amongst the remaining ones is chosen to assemble the population $S^{t'}$ with N individuals. S^t and $S^{t'}$ form the mating pool. Next, the best individual s^* in the current population S^t, denoted by *best*, is chosen. For instance, for minimization problems *best* is such that

$$best = argmin_{s \in S^t} f(s). \tag{17}$$

Selection chooses, by looking at each individual of S^t and the corresponding mate in $S^{t'}$, the one which is the closest to *best*. Recombination is done pairwise between the individuals of the mating pool, weighted by their values of compatibility and arousal. Mutation uses the selected and recombined individuals to produce variations whose amount is weighted by compatibility and arousal as well. If a offspring is better than the current best individual, then it replaces the current *best*. Otherwise, if a mutated individual is better than current best individual, then it replaces the

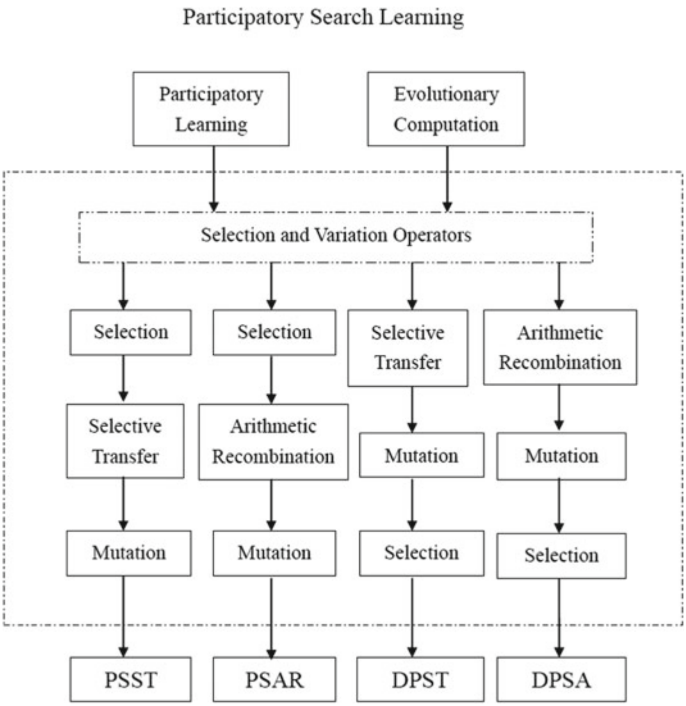


Fig. 7 Participatory search algorithms

current *best*. A new iteration starts with a new population S^{t+1} composed by the current best individual, with the remaining $(N-1)$ individuals chosen randomly. We should remark that participatory search algorithms are elitist: the best individual encountered is always kept in a population. The directive $last(S^t) \leftarrow best$ means that the best individual found up to generation t, denoted by *best*, is kept at the position that corresponds to the last individual of the population at step $t+1$.

There are four instances of PSA, respectively, participatory search with selective transfer (PSST), participatory search with arithmetic recombination (PSAR), differential participatory search with selective transfer (DPST), and differential participatory search with arithmetic recombination (DPSA). They are distinguished by the nature of the recombination, and the order in which the operations of selection, recombination, and mutation are processed in each generation. They also differ from similar evolutionary approaches developed in [6, 7, 18] in the way the mating pool is constructed to produce the new population. A class of participatory search algorithms that incorporates participatory learning is shown in Fig. 7.

PSST is similar to the algorithm discussed in [6] in the sense that both algorithms use participatory selective transfer and mutation. PSAR uses participatory arithmetic recombination and mutation, processed in a different order than PSST. DPST is similar to the algorithm of [7] because it also uses selective transfer and participatory mutation. Likewise, DPSA is similar to the algorithm of [18] and uses participatory arithmetic recombination and mutation. DPSA proceeds similarly as DPST except that it uses arithmetic recombination instead of selective transfer. PSST, PSAR, DPST and DPSA differ from all previous approaches because selection is done individually for each of the N individuals of the current population. Participatory recombination and mutation are performed likewise. Recall that PSST, PSAR, DPST and DPSA are all elitist: the best individual is always kept in the current population. As an illustration, the procedure PSAR is detailed below. The remaining algorithms, except for their nature, have similar format. A in-depth description, characterization, and convergence analysis of the PSA can found in [19].

---

procedure PSAR
  f an objective function
  $s \in S^t$ and $s' \in S^{t'}$
  set *best* randomly
  set $a_0 \leftarrow 0$; $t \leftarrow 0$
  while $t \leq t_{max}$ **do**
    generate population S^t randomly
    $last(S^t) \leftarrow best$
    $S^{t'} \leftarrow s' = argmax_{r \in S^t}(\rho(s, r))$
    find *best* in S^t
Selection:
    compute $\rho^s(s, best)$ and $\rho^{s'}(s', best)$
    if $\rho^s \geq \rho^{s'}$ **then**
      $p_{selected} \leftarrow s$
    else
      $p_{selected} \leftarrow s'$
    end if
Recombination:
    choose $\alpha, \beta \in [0, 1]$ randomly
    compute $\rho_r = \rho(s, s')$
    compute $a_{t+1} = a_t + \beta((1 - \rho_r) - a_t)$
    $p_r = (1 - \alpha\rho_r^{1-a_t})s + \alpha\rho_r^{1-a_t}s'$
Mutation:
    compute $\rho_m = \rho(p_{selected}, p_r)$
    $p_m = best + \rho_m^{1-a_{t+1}}(p_{selected} - p_r)$
    if $f(p_r)$ better than $f(best)$ **then**
      $best \leftarrow p_r$
    end if
    if $f(p_m)$ better than $f(best)$ **then**
      $best \leftarrow p_m$
    end if
    $t \leftarrow t + 1$
  end while
  return *best*
end procedure

---

6 Participatory Search Algorithms in Fuzzy Modeling

This section concerns the use of participatory search algorithms in fuzzy rule-based system modeling. The aim is to illustrate potential applications of PSA and to evaluate and compare the performance of PSST, PSAR, DPST and DPSA algorithms using actual data and results reported in the literature.

The problem of interest here is to develop linguistic fuzzy models using actual data sets available in KEEL (http://www.keel.es/). The KEEL (Knowledge Extraction based on Evolutionary Learning) is a software tool to assess evolutionary algorithms for data mining problems including regression, classification, clustering, and pattern mining. KEEL provides a complete set of statistical procedures for multiple comparisons. The features of the data sets are summarized in Table 1. These data are the same used in [1], a state of the art representative GFS reported in the literature [20]. The representation and encoding schemes of PSAR are also the same of the one adopted in [1]. They are as follows:

1. Database encoding: ($C = C_1, C_2$) a double-encoding scheme.
   First, equidistant strong fuzzy partitions are identified considering the granularity (labels) specified in C_1. Second, the membership functions of each variable are uniformly rearranged to a new position considering lateral displacement values specified in C_2.

   - Number of labels C_1: this is a vector of integers of size n representing the number of linguistic variables.

   $$C_1 = (L^1, ..., L^n). \tag{18}$$

     Gene L^i is the number of labels of the ith linguistic variable, $L^i \in \{2, ..., 7\}$.
   - Lateral displacements C_2: this is a vector of real numbers of size n that encodes displacements α^i of the different variables, $\alpha^i \in [-0.1, 0.1]$. A detailed description of the linguistic 2-tuple representation is given in [21, 22].

Table 1 Summary of the datasets

Problem	Abbr.	Variables	Samples
Electrical maintenance	ELE	4	1056
Auto MPG6	MPG6	5	398
Analact	ANA	7	4052
Abalone	ABA	8	4177
Stock prices	STK	9	950
Forest fires	FOR	12	517
Treasury	TRE	15	1049
Baseball salaries	BAS	16	337

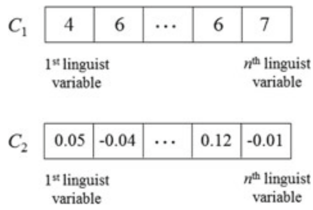


Fig. 8 A double-encoding scheme C_1 and C_2

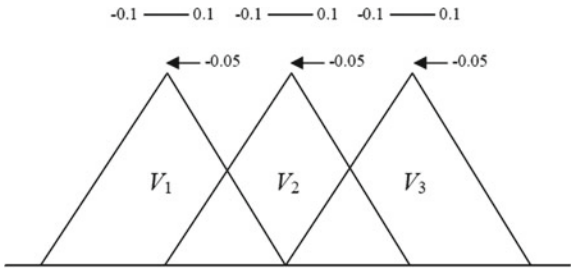


Fig. 9 Lateral displacement of the linguistic variable V values V_1, V_2, and V_3

$$C_2 = (\alpha^1, ..., \alpha^n). \tag{19}$$

An example of the encoding scheme is given in Fig. 8.
Figure 9 illustrates the lateral displacement of V for $\alpha = -0.05$.

2. Rule base: constructed using the Wang and Mendel algorithm (WM) [23, 24] as follows:

   a. granulate the input and output spaces;
   b. generate fuzzy rules using the given data;
   c. assign a certainty degree to each rule generated to resolve conflicts;
   d. create a fuzzy rule base combining the rules generated and rules provided by experts (if available);
   e. determine the input-output mapping using the combined fuzzy rule base and a defuzzification procedure.

   An example of a fuzzy rule-base developed for ELE is shown in Fig. 10.
   Example of rules of the rule base of Fig. 10 include:

rule 1: IF X1 is 1 and X2 is 1 and x3 is 1 and x4 is 1 THEN Y is 1
rule 2: IF X1 is 2 and X2 is 1 and x3 is 1 and x4 is 2 THEN Y is 3
rule 3: IF X1 is 3 and X2 is 3 and x3 is 2 and x4 is 3 THEN Y is 4

3. Objective function: the mean-squared error (*MSE*)

$$MSE = \frac{1}{2|D|}\sum_{l=1}^{|D|}(F(x^l) - y^l)^2 \tag{20}$$

Fig. 10 Rule base constructed using WM algorithm

 # Rule base : 28

X1	X2	X3	X4	Y
1	1	1	1	1
2	1	1	2	3
3	3	2	3	4
3	2	1	3	3
3	4	3	2	4
1	2	1	1	1
1	1	1	2	2
1	2	1	2	2
2	2	2	3	3
4	2	1	2	3
..	..	..	..	..

where $|D|$ is the size of the dataset, $F(x)$ is the output of the FRBS model, and y the actual value of the output. Fuzzy inference uses the max-min procedure with center of gravity deffuzification.

4. Initial population: each chromosome has the same number of linguistic labels, from two to seven labels for each input variable. For each label of the inputs, all possible combinations are assigned to the respective rules consequents. Moreover, for each combination, two copies are added with different values in the C_2 part. The first has values randomly chosen in $[-0.1, 0]$ and the second random values chosen in $[0, 0.1]$.
5. Recombination: $p_r \leftarrow floor(p_r)$ for C_1.
   If a gene g of p_r in C_1 is lower than 2, then $L_g = 2$, else if a gene g is higher than 7, then $L_g = 7$.
6. Mutation: $p_m \leftarrow floor(p_m)$ for C_1.
   If a gene g of p_m in C_1 is lower than 2, then $L_g = 2$, else if a gene g is higher than 7, then $L_g = 7$.

The electric maintenance model has four input variables and one output variable. The ELE dataset contains electrical maintenance data and has 1056 samples. This is an instance in which we expect learning methods to develop large number of rules. ELE modeling involves a large search space [1]. The MPG data concerns city-cycle fuel consumption in miles per gallon (mpg), to be predicted in terms of one multivalued discrete and five continuous attributes. The MPG6 dataset has 398 samples. The categorical data (ANA) is one of the data sets used in the book *Analyzing Categorical Data* by Jeffrey S. Simonoff. It contains information about the decisions taken by a supreme court. The ANA dataset concerns seven variables and 4052 samples. Abalone age data come from physical measurements. The abalone model has eight input variables and one output variable. The abalone dataset (ABA) contains 4177 samples. The STK data provided are daily stock prices from January 1988 through October 1991, for ten aerospace companies. The task is to approximate the price

Table 2 Methods considered by the computational experiments [1]

Method	Type of learning
WM(3)	Rule base produced by WM, 3 linguistic labels for each variable
WM(5)	Rule base produced by WM, 5 labels for each variable
WM(7)	Rule base produced by WM, 7 labels for each variable
FSMOGFS	Gr. Lateral partition parameters, and rule base produced by WM
FSMOGFS+TUN	FSMOGFS + Tuning of MF parameters and rule selection by SPEA2
FSMOGFS^e+TUN^e	FSMOGFS+TUN including fast error estimation

of the 10th company given the prices of the rest. The STK has nine input variables and 950 samples. The FOR dataset has 12 variables and 517 samples. The aim is to predict the burned area of forest fires, in the northeast region of Portugal. The TRE contains the economic data information of USA and has 15 variables input and 1049 samples. The goal is to predict 1-Month Rate. The BAS contains the salaries of the set of Major League Baseball players and has 16 variables input and 337 samples. The task is to approximate the salary of each player. The datasets are available at http://sci2s.ugr.es/keel/index.php. The methods considered in [1] are summarized in Table 2. The method of Wang and Mendel (WM) is also a reference because all PSA and the GFS use it as a rule generation procedure during evolution. The participatory search algorithms were run using the datasets to compare their results with the ones produced by PSAR and results reported in the literature [1]. The processing times of the different methods in [1] were obtained using an Intel Core 2 Quad Q9550 2.83-GHz, 8 GB RAM. The processing times of participatory search algorithms reported here were obtained using an Intel Core 2 Quad Q8400 2.66GHz, 4 GB RAM.

The input parameters used by participatory search algorithms in the experiments reported in this section are: population size of 60, and maximum number of function evaluations of 1000. Data sets were randomly split into five folds, each partition containing 20% of the dataset. Four of these partitions are used for training and the remaining one is used for testing. The algorithms are run six times for each data partition using six distinct seeds.

The results show that the average mean-squared error for the test data achieved by the fuzzy models developed by PSAR, Table 6, is lower than the average mean-squared error of test data achieved by the FSMOGFS^e+TUN^e, except for ANA data. Also, the average mean-squared error for the test data achieved by DPSA is lower than the FSMOGFS^e+TUN^e. For the test data of ANA, FSMOGFS^e+TUN^e achieves the lowest *MSE* value. Considering the test data PSAR, with WM using different number of labels for each linguistic variable, is more accurate than when the number of linguistic labels for each linguistic variable is kept fixed, WM(3),WM(5) and WM(7), respectively. Thus, PSAR performs better than FSMOGFS^e+TUN^e from the point of view of the test data of *MSE*. Also, the standard deviation (SD) of test data for PSAR and FSMOGFS^e+TUN^e is better than WM(3),WM(5) and WM(7).

Table 3 Average rank of the algorithms

Algorithm	Friedman rank	p-value	H_0
WM(3)	7.3125		
WM(5)	6.25		
WM(7)	6		
FSMOGFS^e+TUN^e	3.75	1.38E-7	Rejected
PSAR	2.125		
PSST	4		
DPSA	2.3125		
DPST	4.25		

Table 4 Holm's Post-Hoc for $\varepsilon = 0.05$.

Control algorithm: PSAR					
i	Algorithm	z value	p-value	ε/i	H_0
7	WM(3)	4.2355	2.3E-5	0.0071	Rejected
6	WM(5)	3.368	0.0007	0.0083	Rejected
5	WM(7)	3.1639	0.0015	0.01	Rejected
4	DPST	1.735	0.08273	0.0125	Rejected
3	PSST	1.5309	0.1257	0.0166	Rejected
2	FSMOGFS^e+TUN^e	1.3268	0.184573	0.025	Accepted
1	DPSA	0.153	0.8783	0.05	Accepted

Further analysis is pursued as suggested in [25, 26] to verify if there exist statistical differences among the performance of the algorithms. Recall that the confidence level is $\varepsilon = 0.05$. Table 3 shows how PSAR and GFS are ranked. PSAR achieves the highest rank with 1.375. Also, recall that the null hypothesis H_0 is that PSAR and GFS algorithms are equivalent, that is, H_0 means that the rank of all algorithms are equal. If the hypothesis is rejected, then we conclude that the algorithms perform differently.

Iman-Davenport's test suggests that there are significant differences among the algorithms in all datasets once the null hypothesis is rejected (p-value $= 1.38\,E{-}7$). Thus the Holm post-hoc test is conducted with PSAR as the control algorithm. Table 4 shows that the Holm post-hoc test rejects the hypothesis concerning WM(3), WM(5), WM(7), DPST and PSST, but do not reject FSMOGFS^e+TUN^e and DPSA. Therefore, PSAR outperforms WM(3), WM(5), WM(7), DPST and PSST because the rank pf PSAR is the highest and rejects the hypothesis in the Holm test. We notice that the difference of the performance of FSMOGFS^e+TUN^e and DPSA is not statistically relevant because the null hypothesis is accepted.

Table 5 highlights, that for each dataset, the average processing time of FSMOGFS^e+TUN^e and PSAR in minutes and seconds. We notice the different complexity of the solutions generated during the evolutionary process. The com-

Table 5 Average runtime of the algorithms (minutes:seconds M:S)

Dataset	FSMOGFS^e+TUN^e	PSAR
ELE	00:42	00:45
MPG6	1:00	00:53
ANA	5:17	5:05
ABA	3:54	4:25
STK	1:31	1:12
FOR	1:07	00:40
TRE	00:46	1:02
BAS	00:58	1:01

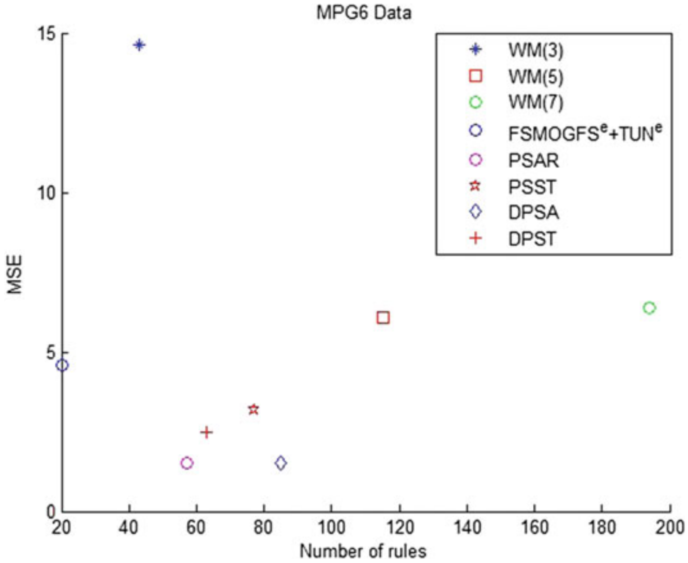


Fig. 11 MSE performance of the algorithms versus the number of rules: MPG6 data

putational cost of the fitness evaluation depends of the number of rules and conditions in rules antecedents. In the case of ANA, the PSAR needs less time than FSMOGFS^e+TUN^e because the number of rules is small. On the other hand, it is higher than 3 min in ANA and ABA because the large number of samples.

In sum, the performance of PSAR in developing fuzzy rule-based models with actual data illustrates its potential to solve complex problems. Overall, the results suggest that PSAR performs better than current state of the art genetic fuzzy system approaches from the point of view of the average mean square error with test data. Figure 11 summarizes the MSE performance of the algorithms versus the number of rules for MPG6 data set. More importantly, participatory search algorithms are

Table 6 Average *MSE* of PSA and GFS Algorithms

		WM(3)		WM(5)		WM(7)		FSMOGFS^e+TUN^e		PSAR		PSST		DPSA		DPST	
Dataset		Tra	Tst	Tra	Tst	Tra	Tst	Tra	Tst	Tra	Tst	Tra	Tst	Tra	Tst	Tra	Tst
ELE	Rule	27		65		103		8		28		27		23		23	
	Mean	192241	192647	56135	56359	53092	55495	9665	10548	9502	10480	11409	10560	9586	10434	11250	10544
	SD	9658	14436	1498	4685	1955	9452	823	1150	3951.94	3986.8	3874.74	3906.8	3914.4058	3753.2432	3020.37	3260.55
MPG6	Rule	43		115		194		20		57		77		85		63	
	Mean	13.552	14.649	4.136	6.096	2.642	6.382	2.86	4.562	1.6132	1.5205	3.424	3.1699	1.8901	1.5151	2.0641	2.467
	SD	1.239	3.204	0.317	2.416	0.11	2.126	0.218	0.714	1.3712	1.2983	1.3093	1.2639	1.1779	1.2563	1.1532	1.299
ANA	Rule	72		124		171		10		6		4		7		5	
	Mean	0.187	0.189	0.027	0.03	0.012	0.017	0.003	0.003	0.0256	0.0856	0.0292	0.0682	0.0274	0.0507	0.0292	0.0795
	SD	0.001	0.005	0	0.002	0	0.003	0	0.001	0.0833	0.0574	0.0875	0.0485	0.0913	0.0866	0.0836	0.0373
ABA	Rule	68		199		368		8		34		23		34		30	
	Mean	8.407	8.422	3.341	3.474	3.057	3.268	2.445	2.509	0.0016	0.0047	0.0018	0.005	0.0018	0.0047	0.0055	0.0048
	SD	0.443	0.545	0.13	0.247	0.084	0.185	0.114	0.184	0.0002	0.0002	0.0001	0.0002	0.0002	0.0002	0.0005	0.0002
STK	Rule	123		265		378		23		58		73		66		78	
	Mean	8.852	8.951	1.576	1.624	0.611	1.488	0.764	0.912	0.7342	0.9002	1.1022	0.9064	0.7784	0.8139	1.6701	1.8139
	SD	0.508	1.193	0.09	0.09	0.029	1.634	0.139	0.181	0.0479	0.1001	0.0687	0.1333	0.0912	0.0597	0.0738	0.0987
FOR	Rule	246		375		401		10		23		31		8		43	
	Mean	2030	3793	1435	34235	340	1.00E+05	1418	2628	2115.76	1619.13	2231.0333	2.35E+03	2652.8666	2.04E+03	1600.9333	1.74E+03
	SD	531	2340	505	4356	147	4708	539	2108	381.98	756.86	169.3108	450.3585	207.1127	461.7959	194.4706	524.3924
TRE	Rule	75		196		261		9		26		28		30		25	
	Mean	1.636	1.631	0.401	0.405	0.17	0.176	0.034	0.044	0.0529	0.0372	0.1168	0.1707	0.0655	0.0392	0.1492	0.128
	SD	0.121	0.181	0.014	0.055	0.009	0.017	0.003	0.015	0.0059	0.0045	0.0059	0.019	0.0067	0.008	0.0099	0.0205
BAS	Rule	181		253		264		17		50		52		47		65	
	Mean	1.9E+05	3.6E+05	7.8E+04	6.2E+05	3.1E+04	1.0E+06	1.4E+05	2.6E+05	1.5E+05	1.3E+05	1.0E+05	1.5E+05	1.1E+05	3.6E+05	1.0E+05	3.9E+05
	SD	1.0E+04	7.3E+04	4.7E+03	6.8E+04	6.0E+02	1.3E+05	1.9E+04	5.8E+04	7753.67	124393	5.9E+03	3.1E+04	7.6E+03	2.6E+05	7.2E+03	3.0E+05

simpler, have high computational performance, and require few parameters to run. In particular, PSAR is a highly competitive population-based search approach (Table 6).

7 Conclusion

Participatory search is a population-based instance of the participatory learning paradigm. Compatibility degrees and arousal indexes account for the effect of the population individuals during search. Recombination arises from an instance of participatory learning formula. The participatory search algorithms are elitist and employ compatibility and arousal information in selection, recombination and mutation. Applications concerned the use of participatory search algorithms to develop fuzzy linguistic models of actual data. The performance of the models produced by the participatory search algorithms were evaluated and compared with a state of the art genetic fuzzy system approach. The results suggest that the participatory search algorithm with arithmetical-like recombination performs best.

Acknowledgements The second author is grateful to CNPq, the Brazilian National Council for Scientific and Technological Development (CNPq), for grant 305906/2014-3.

References

1. Alcalá, R., Gacto, M., Herrera, F.: A fast and scalable multiobjective genetic fuzzy system for linguistic fuzzy modeling in high-dimensional regression problems. IEEE Trans. Fuzzy Syst. **19**(4), 666–681 (2011)
2. Herrera, F., Verdegay, J.: Genetic Algorithms and Soft Computing. Physica-Verlag, Heidelberg, Germany (1996)
3. Storn, R., Price, K.: Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. **11**, 341–359 (1997)
4. Yager, R.: Participatory Genetic Algorithms. BISC Group List, message posted on 29 Aug 2000
5. Yager, R.: A model of participatory learning. IEEE Trans. Syst. Man Cybern. **20**(5), 1229–1234 (1990)
6. Liu, Y.L., Gomide, F.: Participatory genetic learning in fuzzy system modeling. In: Proceedings of IEEE SSCI 2013, Singapore (2013)
7. Liu, Y.L., Gomide, F.: Evolutionary participatory learning in fuzzy system modeling. In: Proceedings of IEEE International Conference on Fuzzy System, p. 2013, Hyderabad, India (2013)
8. Liu, Y.L., Gomide, F.: Participatory search algorithms in fuzzy modeling.In: Proceedings of the World Conference in Soft Computing, Berkeley, USA (2016)
9. Ishibuchi, H., Narukawa, K., Tsukamoto, N., Nojima, Y.: An empirical study on similarity-based mating for evolutionary multiobjective combinatorial optimization. Eur. J. Oper. Res. **188**, 57–75 (2008)
10. Fogel, D.: An Introduction to Evolutionary Computation, Chapter 1 of Evolutionary Computation: The Fossil Record. Wiley-IEEE Press, New York, USA (1998)
11. Glover, F., Marti, R.: Fundamentals of scatter search and path relinking. Control Cybern. **29**(3), 653–684 (2000)

12. Voget, S., Kolonko, M.: Multidimensional optimization with a fuzzy genetic algorithm. J. Heuristics **4**(3), 221–244 (1998)
13. Hwang, H.-S.: Control strategy for optimal compromise between trip time and energy consumption in a high-speed railway. IEEE Trans. Syst. **28**(6), 791–802 (1998)
14. Cordón, Ó.: Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases. In: Advances in Fuzzy Systems. World Scientific Publishing, Singapore (2001)
15. Brown, R.: Moothing, Forecasting and Prediction of Discrete Time Series. Prentice-Hall, New Jersey, USA (2004)
16. Birchenhall, C., Lin, J.-s.: Learning and adaptive artificial agents: Analysis of an evolutionary economic model. In: Computing in Economics and Finance, vol. 327. University of Manchester, United Kingdom (2000)
17. Hwang, H.-S.: Genetic algorithms in evolutionary modelling. J. Evolut. Econ. **7**(4), 375–393 (1997)
18. Liu, Y.L., Gomide, F.: Fuzzy systems modeling with participatory evolution. In: IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS). Edmonton, AB, Canada (2013)
19. Liu, Y.L.: Participatory search algorithms and applications. Ph.D. Thesis, School of Electrical and Computer Engineering, University of Campinas, Campinas, Sao Paulo, Brazil (2016)
20. Fazzolari, M., Alcala, R., Nojima, Y., Ishibuchi, H., Herrera, F.: A review of the application of multiobjective evolutionary fuzzy systems: current status and further directions. IEEE Trans. Fuzzy Syst. **21**, 45–65 (2013)
21. Herrera, F., Martínez, L.: A 2-tuple fuzzy linguistic representation model for computing with words. IEEE Trans. Fuzzy Syst. **8**(6), 746–752 (2000)
22. Alcalá, R., Alcalá-Fdez, J., Herrera, F., Otero, J.: Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation. Int. J. Approx. Reason. **44**(1), 45–64 (2007)
23. Wang, L., Mendel, J.: Generation fuzzy rules by learning from examples. IEEE Trans. Syst. Man Cybern. **22**(6), 1414–1427 (1992)
24. Wang, L.: Adaptive Fuzzy Systems and Control: Design and Stability Analysis. Prentice-Hall, Upper Saddle River, USA (1994)
25. Derrac, J., García, S., Molina, D., Herrera, F.: A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. Swarm Evolut. Comput. **1**, 3–18 (2011)
26. Antonelli, M., Ducange, P., Marcelloni, F.: An efficient multi-objective evolutionary fuzzy system for regression problems. Int. J. Approx. Reason. **54**, 1434–1451 (2013)

But, What Is It Actually a Fuzzy Set?

Enric Trillas

Abstract Supported in a new view of meaning as a quantity in whatever universe of discourse, and for its possible use concerning plain language and ordinary reasoning in 'Computing with Words', the paper deals with the basic concept of a fuzzy set. That is, not only with the collective a linguistic label generates in language, but also with what membership functions reflect on it once ideally seen as measures of meaning.

1 Introduction

More than fifty years after its introduction [1], the idea of a fuzzy set is not yet clear enough, and although no 'if and only if' definition of it exists, too often fuzzy sets are seen as if they were mathematical entities in themselves instead of linguistic entities.

The identification of a fuzzy set with a single one of its possible membership functions is something very bizarre, since the question that should be immediately posed is, 'But, which one of them?'. Paraphrasing Quine's words [2], it does not seem possible to describe an entity without some criteria of identity; and the correspondence between fuzzy sets and membership functions is one-to-many.

In addition, identifying the fuzzy sets in a universe of discourse X with all the functions in $[0, 1]^X$ is still another oddity, since no general criteria are known for assigning to each one of these this enormous amount of functions a linguistic label of which it can be a membership function. In front of the unknown multiplicity of predicative words acting on X, it is the aleph-one cardinality of the functions in $[0, 1]^X$. Furthermore, an important characteristic usually required of membership

To Professor José Luis Verdegay, with deep affection for his academic work. Curro, thanks for the large friendship we jointly kept from so long ago!

E. Trillas (✉)
University of Oviedo, Asturias, Spain
e-mail: etrillasetrillas@gmail.com

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_12

functions is their continuity, for which a topology in X is necessary; when X can be seen as a subset of the real line such topology is automatically given, but, in general its existence is unknown. In plain language, words do act in whatsoever universe of discourse.

Nevertheless, most books on fuzzy set theory and fuzzy logic usually begin under these presumptions that, although sometimes not explicit but implicitly present in them, not even can be sustained as simple ‘working hypothesis’. Hence both theories usually appear as based on moving grounds; trying to find, at least, a more solid and suitable basement for the idea of fuzzy set is the only goal of this paper.

2 Language and Fuzzy Sets

2.1 What does not seem to be debatable is that a fuzzy set in a universe of discourse X comes from a linguistic label, that is, a predicative word P whose behavior in X is manifested through the elemental statements ′x is P′, and generates the ‘fuzzy set labeled P′. Each one of these statements reflects that a property p, with name P, not only holds up to some extent for the elements in X, but is exhibited by them and is externally recognized.

For instance, in the universe of the London inhabitants, the word ‘young’ is so well anchored in plain language that it allows to speak on the ‘young Londoners’; something that is understood by everybody. In the universe of the positive integers, the word ‘odd’ is also so well anchored in the language of Arithmetic, that it allows to speak on the ‘odd numbers’. In a universe of trains, the word ‘large’ is so well anchored in plain language that it allows to speak on the ‘large trains’. At each part of language, its speakers perfectly understand these statements. Often, in plain language, recognizing p in the elements in X has an empirical character.

There are predicative words generating ‘linguistic collectives’, that is, words that ‘collectivize’ in language [3]. When such language is an artificial one, like it is that of Arithmetic, those words whose behavior is defined by ‘if and only if’ conditions, are the precisely used words that, of course, also exist in plain language. For instance, and in a universe of discourse constituted by two people, John and Sarah, the word ‘young’ can be undoubtedly applied to Sarah provided it is known that she is 18 years old, but John is 67. Notwithstanding, the same word ‘young’ does not admit a precise use in a universe with a very big number of elements like it is that of London inhabitants; for instance, provided it were accepted that ′x is young’ if and only if x is no more than thirty years old, is that a Londoner of thirty years and a few days old, should cease to be qualified as young? Obviously, it is not in plain language where its speakers see linguistic collectives as a single entity.

It can be said that the use of precisely used words is rigid, but that of the not precisely used is flexible. Hence, the first are those words that collectivize in just a subset of the universe of discourse; they perfectly classify the universe in those elements fulfilling the property p, and those not fulfilling p. It just corresponds to the 'axiom of specification' in the Naïve Set Theory [4], stating that a bi-valued property p generates in X a perfect classification of its elements in two subsets, that with those x fully verifying p, and that with those not verifying p at all. A representation of a precisely used word P in X, is done by the set **P** stated by the axiom of specification; its use is rigid, and the Boolean algebra of the subsets in X is the non debatable structure for representing these linguistic labels; it is a mathematical model perfectly mimicking how precise words are used when all the necessary information on their use is, at least, potentially known.

The words not precisely used, and as it is shown by a 'Sorites' type argument [5] applied to each one of them, cannot collectivize in a subset. The, well anchored in plain language, linguistic collectives they generate are like gaseous, or cloudy, entities; but, anyway, gas volumes and clouds do actually exist, and are scientifically studied. Indeed, linguistic collectives are rooted in plain language, and founding a mathematical model for them is imperative; at least for symbolically representing the ordinary reasoning that is greatly permeated by imprecision.

2.2 It seems that the study of the collectives generated by not precisely used words deserve to be approached by not considering them as pure mathematical entities characterized by an 'if and only if' definition, but in a scientific-like style. Such entities should be seen in a different way than sets, but, in any case, they can, and will, be identified with the fuzzy sets; that is, naïvely renaming them as fuzzy sets. The fuzzy set in X labeled P is just the linguistic collective P generates in X, the collective named 'the Ps of X′', and it allows to see a fuzzy set as a single, although not precise, entity. In this way, the linguistic collective generated by a precisely used word is just a subset of X; subsets represent but degenerate, bi-valued, collectives; a non-degenerated collective is a purely linguistic concept.

In this line of thought, the fuzzy set labeled P can be seen as a single, although cloudy, linguistic entity rooted in X. But it still lacks to answer the question: How a fuzzy set can be specified in such a way that the axiom of specification for the precisely used words can follow from it? A possible answer for this question lies in 'meaning'; it is a semantic topic.

3 Words with Measurable Meaning

3.1 The meaning of a word is not independent of the context of its use; for instance, 'odd' does not mean the same in a universe of positive integer numbers than in one of people. Sometimes the meaning of words can even change depending on the purpose for its use, as it is with 'odd' when used to qualify people either with a joking, or a hilarious, or an insulting purpose. It suffices to look for words in a dictionary to check all this by seeing how their uses are described.

Hence, no realistic theory of meaning can assign to words a single and universal meaning, since it depends on the universe of discourse, and on the particular context of its use; the meaning of words is context-dependent and purpose-driven. Thus, and in addition, for arriving at a theory of meaning it should be considered how P ‘behaves’, or ‘acts’ in X. Language is not static, but dynamic; almost always, time intervenes in language.

Once a pair (X, P) is given, how the ‘behavior’ of the word P in the universe X can be described? It only can come from considering the action of P for the x in X; that is, from the elemental statements ′x is P′, that constitute a different entity X [P] than X; the second can be physical or virtual, but the first is always virtual. Nevertheless, to capture the ‘behavior’ of P in X, it should be recognized not only the action of P on the elements x, but the context on which the statements ′x is P′ are used. To capture how P behaves, or acts, in X, it not suffices to statically capture what indicate the statements ′x is P′, it is also necessary to know how such action varies along X; its internal dynamism. That is, recognizing the linguistic relationship [6],

$$'\text{x is less P than y}', \text{ equivalent to } '\text{y is more P than x'};$$

in other words, that x verifies p, less than y does. Such recognition is, in the case of plain language, often of an empirical character not immediately allowing the assignation of a degree to the verification of p.

Let’s, symbolically, denote it by $x <_P y$, and by $x =_P y$ the case in which both $x <_P y$ and $y <_P x$, or $x <_P^{-1} y$, hold; obviously, $=_P \; = \; <_P \cap <_P^{-1}$.

To avoid the possibility $<_P = \emptyset$, let’s state that

$$x <_P x \text{ holds for all x in X, that } <_P \text{ is a reflexive relation},$$

although no reason for stating other properties like symmetry, or transitivity, etc., can really, and generally, exist. In general, $<_P$ is not a partial order in X; less again it is a total, or linear, relation since usually non comparable elements will exist, that is, pairs x, y for which it is neither $x <_P y$, nor $y <_P x$.

In the case the word P is precisely used in X, the relation $<_P$ collapses in the relation $=_P$; that is, $<_P \; = \; <_P \cap <_P^{-1} \iff <_P \; = \; <_P^{-1}$. For instance, there is no way to directly stating that ‘7 is less odd than 517′; all odd numbers are equally odd; analogously, no way of directly establishing that ‘7 is less prime than 17′ exists, all prime numbers are equally prime. The word ‘directly’ refers to doing it without a previous new definition of ‘less than’ for the corresponding word, but only under the old definitions of ‘odd’ and ‘prime’. Numbers are perfectly classified in ‘odd’ and ‘not odd’, ‘prime’ and ‘not prime’, etc.

Once the relation $<_P$ is known, the graph $(X, <_P)$ reflects the ‘semantic organization’ the use of P introduces in X; the graph is a *qualitative meaning* of P in X. It should be pointed out that the universe X cannot be always supposed to be mathematically structured; in plain language, words act in universes whatsoever. For instance, the Kolmogorov’s theory of probability concerns ‘probable’ events in a Boolean algebra, but, in plain language, the same word ‘probable’ is not only applied to such kind of rigid events. A theory of linguistic meaning cannot presume that the universe of discourse is directly endowed with a particular mathematical structure.

Notice that such a definition remembers the intuitive idea that, when (intelligently) talking on some subject, some 'ordering' between the statements that are uttered, gestured, or written at the respect and for the corresponding reasoning's argumentation, is tried to be introduced among them. Additionally, the former definition also seems to be in agreement with the famous Ludwig Wittgenstein's statement [7], 'The meaning of a word is its use in language'.

3.2. Once a graph $(X, <_P)$ is known, the possibility of measuring to which extent x verifies p is open. A measure on such graph is a mapping m_P: $X \rightarrow [0, 1]$, such that [8]:

(1) If $x <_P y$, then $m_P(x) \leq m_P(y)$,
(2) If x is maximal in the graph, that is, no other y verifying $x <_P y$ does exist, then $m_P(x) = 1$,
(3) If y is minimal in the graph, that is, no other x verifying $x <_P y$ does exist, then $m_P(x) = 0$.

Notice that the closed unit interval could be substituted by any closed interval [a, b] in the real line, by just a playing the character of 0, and b that of 1 in the former definition. Notice also that no additive law is presumed for m_P; its definition is free from considering 'and', 'or', and either the concepts of incompatibility or contradiction that are only indistinguishable in the framework of Boolean algebras.

The additive law is deeply involved with a 'rigid form' of classifying elements, and, in plain language contradiction is independent of incompatibility, contrarily, for instance to the case of Ortho-lattices where the first implies the second ($p \leq q' \Rightarrow p \cdot q = 0$), with the reciprocal only holding provided the Ortho-lattice is a Boolean algebra, that is, it is distributive and consequently verifies the law of 'perfect repartition', $p = p \cdot q + p \cdot q'$, for all pair p, q. In this case, $p \cdot q = 0$ implies $p = p \cdot q' \Leftrightarrow p \leq q'$, and the reciprocal also holds. A lot of structural laws is necessary for the equivalence between contradiction and incompatibility; something that cannot be presumed in plain language.

The former general definition of a measure is inspired on that of a 'fuzzy measure' introduced by Michio Sugeno [9], but liberated from the constraints imposed by just measuring subsets; it is free from any mathematical structure in X further than that of graph. An antecedent of it can be found in the concept of a 'fuzzy entropy', introduced by Aldo de Luca, and Settimo Termini [10], where P = fuzzy and $<_{\text{fuzzy}}$ coincides with the so called 'sharpened order' between functions in $[0, 1]^X$.

Like, for instance, in the case with probabilities, with Sugeno's λ-measures, or with de Luca-Termini fuzzy entropies, all of them measures, the three axioms of a measure are not sufficient for individuating a single m_P, and additional suppositions should be added for each one of them. Anyway, each triplet $(X, <_P, m_P)$ facilitates a quantity reflecting a *quantitative meaning* of P in X, and, in this way, each 'full meaning' can be seen as a quantity.

By paraphrasing Lord Kelvin's words [11], 'If you cannot measure it, it is not Science', viewing meaning as a quantity can open the door towards a scientific-like study of collectives. Let's show a toy-example to illustrate what has been said.

Consider X = [0, 10], and P = big, generating the linguistic collective, or fuzzy set 'big numbers between 0 and 10′. This collective is not a set, provided 'big' is not rigidly but flexibly used; for instance, provided it can be stated that '8 is big', also '7.99 is big' can be stated. A 'Sorites' argument [5] shows that, not being the 'collective big' empty since 10 is always considered big, it is not a subset of [0, 10]. The qualitative use of big in [0, 10] can be described by:

$$x <_{big} y \Leftrightarrow x \leq y,$$

that is, by the qualitative meaning ([0, 10], $\leq$), a graph that is but a linearly ordered interval with maximum 10 (the only maximal), and minimum 0 (the only minimal). Hence, the measures of big in [0, 10] are the mappings m_{big}: [0, 10] $\rightarrow$ [0, 1], such that,

(1) $x \leq y => m_{big}(x) \leq m_{big}(y)$
(2) $m_{big}(10) = 1$
(3) $m_{big}(0) = 0$,

to which, adding the condition of 'usual flexibility',

(4) If x can be qualified as big, it exists ε (x) > 0 such that those y in the interval (x – ε (x), x] can be also qualified as big,

that could be translated into m_{big}, as
(4*) m_{big} is continuous in [0, 10].

Hence, the measures of big are the mappings between [0, 10] and [0, 1] that are strictly non-decreasing, and verify the border conditions (2) and (3). There is an enormous amount of them.

Consequently, to specify a measure for the meaning of big, it is necessary to add some additional information on the behavior of big, like it can be on its shape. For instance, provided it is known, or can be reasonably presumed, that the measure should be linear, $m_{big}(x) = ax + b$, it follows that the only possible linear measure is x/10, but provided the information on its shape were that it is quadratic, $m_{big}(x) = ax^2 + bx + c$, several possibilities for the values of a and b are possible, since by (3) it follows c = 0, and from (2) that 100 a + 10 b = 1; for instance, a quadratic measure is $x^2/100$ (with a = 1/100 and b = 0), but, obviously, it is not the only quadratic measure that is possible for 'big'.

In conclusion, the graphs ([0, 10], $\leq$, m_{big}), with m_{big} verifying the axioms (1)–(4), plus some additional information or additional reasonable hypotheses, are the quantities that specify a full meaning of 'big' in [0, 10]. These quantities require the 'design' of the corresponding measures.

Notice that axioms (4) and (4*) exclude to interpret the use of big as a precise one, since it will be specified by a subset of [0, 10] whose measure cannot be continuous but with jumps; notwithstanding, such rigid interpretation is possible by avoiding (4), by renouncing to flexibility. That is by ‘making precise’ the meaning of ‘big’; something that means, indeed, changing the ordinary and often usual, use of ‘big’ in plain language.

In some cases, the designer should add to the axioms, some hypotheses, reasonable for the current situation he is faced with, and like it is the former linear hypothesis. In this way, the linguistic collective P generates is qualitatively described by the graph $(X, <_P)$ once it is known, and shows different ‘informational states’ each one given once a quantity $(X, <_P, m_P)$ is specified.

3.3. The former interpretation of meaning as a quantity, actually preserves what has been said for the precisely used words, and it can be proven as follows.

If P is precisely used in X, the graph is $(X, =_P)$, and then if $x =_P y \Leftrightarrow x <_P y$ & $y <_P x$. It implies $m_P(x) \leq m_P(y)$ & $m_P(y) \leq m_P(x)$, or $m_P(x) = m_P(y)$. Thus, only the values 0 and 1 can be taken by the measure, since there are no other elements than the maximal (those verifying P), and the minimal (those not verifying P). Hence, the subset $m_P^{-1}(1)$ consists in the maximal (or prototypical) elements, and is the set **P** specified by P in X; obviously, it is $m_P^{-1}(0) = \mathbf{P}^c$ that contains the minimal or anti-prototypical elements.

4 Membership Functions

Is there any difference between the former measures of the linguistic label ‘big’, and the membership functions assigned to it in any book on Fuzzy Logic? There is no one. Thus, it can be stated that a membership function is, ideally, but a measure of the qualitative meaning of its linguistic label; is a quantitative ‘informational state’ of the fuzzy set.

Nevertheless, as the word ‘ideally’ tries to remark, this statement should be submitted to caution since a membership function is designed with the information available to its designer; an information not always consisting in all the relation ‘less than’, nor in knowing all its maximal and minimal elements but only some of them. There is some similarity with what happens when saying that the probability of obtaining each one of the six faces in throwing a die is 1/6, but without knowing if the die has some imperfection, or it is a tricky one, or the landing surface is not perfectly plane. An ‘ideal’ die is supposed to have a probability of 1/6 for each one of its faces; but often, ideal dice, those who throw them, and the landing surfaces, are not perfect. To say nothing when the die is tricky, for instance, having inside and in front of a face, a very small piece of plumb, or when it is thrown into sand. Throwing a die is a real situation, and often real situations are not ‘ideal’; analogously, the design of a membership function departs from some real situation in plain language. No doubt that the measures are membership functions, but, are membership functions always measures?

In addition, once the designer arrives at a membership function μ_P, a new relation is automatically established in X:

$$x \leq_{\mu} y \Leftrightarrow \mu_P(x) \leq \mu_P(y),$$

that is a linear one and, consequently, not always coincidental with $<_P$. Thus, provided $\leq_{\mu}$ substitutes $<_P$, the new ‘qualitative meaning after design’ is not already the original one. Provided μ_P were a measure, since

$$x <_P y => \mu_P(x) \leq \mu_P(y) \Leftrightarrow x \leq_{\mu} y, \text{ or } <_P \subseteq \leq_{\mu},$$

the design enlarges the qualitative meaning. Hence, the original qualitative meaning could be changed by the larger one coming from design; design could modify meaning. This is certainly risky, since practitioners usually look at the behavior of P in X after counting with a membership function and just through its shape; with it, they can easily appreciate a larger qualitative meaning than the real one.

In conclusion, in most practical cases, the membership function cannot be exactly a measure, but an often unknown approximation to one of them. Thus, to well representing the meaning of a linguistic label, it is important to know when a membership function can be seen as a good enough approximation to a measure. Working with a membership function not well reflecting the meaning of its linguistic label, can conduct to wrong results coming, for instance, from the fact that the membership function actually represents a different linguistic label.

5 Searching for a Definition and a Theorem

Fuzzy Control counts with a theorem stating on which conditions the computed output universally approaches the real one [12], but it lacks a similar theorem for assuming that a membership function truly approaches a measure. That is, a theorem from whose antecedent it can follow that, given a measure m_P of a qualitative meaning $(X, <_P)$ of P in X, a designed membership function μ_P is ‘good enough’ provided it exists a measure m_P such that either

(1) For all $\varepsilon > 0$ is $|m_P(x) - \mu_P(x)| \leq \varepsilon$, for all x in X,

or

(2) It is minimized the function $\text{Sup}_{x \in X} |m_P(x) - \mu_P(x)|$.

Instead of a system of rules and a defuzzification’s method, what here is initially known are the relation $<_P$, or a part of it (both can be seen as a system of rules), and the axioms m_P should verify.

Both (1) and (2) could be considered as suitable definitions for stating that μ_P approaches m_P. Nevertheless, to prove them as a theorem’s conclusion, some reasonable hypotheses, or some additional contextual information on the behavior of P in X, should be taken into account. But it depends on each particular case, and, for example, on how the designer did build up the membership function, that is, on the information available to s/he on the behavior of P in X. It remains an open question that, perhaps, should be posed from a different point of view as it is, for instance, beginning by modifying up to some limits the relation $<_P$, or the measure, or by just considering some type of measures.

I mention such possibility under the conviction that a generally accepted definition concerning the relation a designed membership function should keep with measures, as well as to proving on which conditions it can hold, is an important topic. Of course, were μ_P a measure, both the definition and the theorem are unnecessary. Anyway, and at least, some sufficient condition for knowing if the membership function approaches a measure will be interesting. Up to when something similar will be found, the design of linguistically described systems will continue being done in a not standardized and blind form.

6 Conclusion

This paper is just a first trial to penetrate on what the idea of a fuzzy set is, and on what its description by membership functions means. That the topic is just open, but not fully achieved, is manifested by the final call towards clarifying which membership functions can be actually considered as a good enough approximation to the fuzzy sets informational states.

What does not seem dubious is that fuzzy sets are linguistic, not mathematical, entities, that rooted in plain language belong to its domain, and that their membership functions should come from a process of design. In themselves, fuzzy sets seem to need a scientific-like study instead of a purely mathematical one. In addition, it yet lacks to count with a standardization of the corresponding design's processes for what concerns the approximate character of membership functions.

Since plain language is full of imprecise words, to mathematically mimicking ordinary reasoning, that is, to establish mathematical models for the not fully deductive types of reasoning, it is strictly necessary to consider imprecision, and, hence managing fuzzy sets for its symbolic representation. Thus, clarifying the idea of 'linguistic collectivization' is relevant.

There are still several problems that remain open for counting with good enough theoretical foundations of Zadeh's 'Computing with Words and Perceptions', whose ground lies in plain language and ordinary reasoning. For instance, ambiguity also permeates plain language and no mathematical model for scientifically managing ambiguity is currently known.

Notwithstanding, what can be asserted is that only one kind of fuzzy set actually exists, and that names like 'type-two fuzzy set' only can refer to the range in which the membership function takes its values.

References

1. Zadeh, L.A.: Fuzzy Sets. Inf. Control **8**, 338–353 (1965)
2. Quine, W.V.O.: Ontological Relativity and Other Essays. Columbia University Press, New York (1969)
3. Trillas, E.: On a model for the meaning of Predicates. In: Seising, R. (ed.) Views on Fuzzy Sets and Systems from Different Perspectives, pp. 175–205. Springer, Berlin (2009)
4. Halmos, P.R.: Naive Set Theory. Van Nostrand, New York (1960)
5. Trillas, E., Urtubey, L.A.: Towards the dissolution of the sorites paradox. J. Appl. Soft Comput. **11**(2), 1506–1510 (2011)
6. Trillas, E., Eciolaza, L.: Fuzzy logic. An Introductory Course for Engineers. Springer, Berlin (2015)
7. Wittgenstein, L.: Philosophical Investigations. Basil Blackwell, London (1958)
8. Trillas, E.: On a Model for the meaning of predicates. In: Seising, R. (ed.) Views of Fuzzy Sets and Systems from Different Perspectives pp. 175–2005. Springer, Berlin (2009)
9. Sugeno, M.: Theory of fuzzy integrals and its applications. Ph.D. Tokyo Institute of Technology, Tokyo (1974)
10. de Luca, A., Termini, S.: A definition of a non-probabilistic entropy in the setting of fuzzy sets theory. Inf. Control **20**, 301–312 (1972)
11. Lord Kelvin, W.T.: Lecture on electrical units of measurement; popular lectures and addresses, I/79. Mc Millan, London (1889)
12. Castro, J.L., Delgado, M.: Fuzzy systems with defuzzification are universal approximators. IEEE Trans. Syst. Man Cybern. **26**, 149–152 (1996)

Gradual Numbers and Fuzzy Solutions to Fuzzy Optimization Problems

Didier Dubois and Henri Prade

Abstract This short paper indicates that early examples of fuzzy elements in a fuzzy set, that is, entities that assign elements to membership values, in contrast with fuzzy sets that assign membership values to elements, can be found in papers by Verdegay in the early 1980, following a line of thought opened by Orlovsky. They are so-called fuzzy solutions to fuzzy optimization problems. The notion of fuzzy element, and more specifically gradual number sheds some light on the ambiguous notion of fuzzy number often viewed as generalizing a number while it generalizes intervals. The notion of fuzzy solution is in fact a parameterized solution, in the style of parametric programming. These considerations show the pioneering contributions of Verdegay to the development of fuzzy optimization.

1 Introduction

In the literature of fuzzy optimisation, initiated by Bellman and Zadeh [1], the standard formulation of optimizing an objective function under some rigidly defined constraints is replaced by the search for a solution with maximal membership grade in a fuzzy set defined by intersecting the fuzzy sets of feasible solutions according to several constraints, and the fuzzy sets of good solutions according to one or several objectives. It is clear that under this relaxed form of optimization problem, constraints and objective functions play the same role and because of the use of the minimum operator to aggregate the various fuzzy sets, there is no compensation allowed between local satisfaction degrees. In some sense, as argued in [5], Bellman and Zadeh's formulation is a pioneering generalization of constraint satisfaction problems, understood as flexible or soft constraint satisfaction problems [10] where

D. Dubois(✉) · H. Prade
CNRS-IRIT, Université Paul, Sabatier 118, Route de Narbonne,
31062 Toulouse, Cedex 9, France
e-mail: dubois@irit.fr

H. Prade
e-mail: prade@irit.fr

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_13

the satisfaction of constraints is a matter of degree. This formulation of fuzzy optimization problems, including multiobjective ones, was extensively applied to linear programming, especially by Tanaka et al. [15], Zimmermann [19] and Chanas [3], and further on used by many scholars.

This formulation of fuzzy optimization now belongs to history as almost every possible development of this formulation has been studied. However, what is striking with this approach is that the difference between constraints and objectives is blurred. Once changed into fuzzy sets, they play the same role (of flexible constraints) in the mathematical formulation. Yet quite early, some scholars in the late 1970s such as Orlovsky [13] and in the early 1980s, especially our colleague and friend Curro Verdegay [16, 17] envisaged the fuzzy optimization problem in a different way, acknowledging the specific role of the objective function as opposed to fuzzy constraints, and proposed the idea of fuzzy solution to fuzzy optimisation problems.

More recently, and in a totally different context, the authors of this note pointed out the ambiguity of the terminology "fuzzy number" [7, 8], often viewed as the fuzzy extension of a number while it generalizes intervals, and the questionable understanding of the notion of defuzzification of a fuzzy set that is often supposed to yield a precise element rather than a crisp set. These considerations led us to consider the concept of fuzzy (or yet gradual) element in a fuzzy set, for which the notion of membership function is in some sense taken the other way around.

The aim of this note is to indicate the close connection between Verdegay's proposal of fuzzy solution, and the notion of fuzzy element in a fuzzy set: a fuzzy solution is a fuzzy element in the fuzzy set of feasible solutions to the fuzzy optimization problem.

2 Fuzzy Solutions to Fuzzy Mathematical Programming Problems

At the origin, the introduction of fuzzy sets in optimization problems led to the maximisation of an objective function $f(x)$ under fuzzy constraints on a set $\mathscr{X}$ of potential solutions x [13, 15]. Typically, the objective function is a mapping $x \in \mathbb{R}^n \mapsto f(x) \in \mathbb{R}$, and the constraints form a fuzzy set of feasible solutions in $\mathbb{R}^n$. The fuzzy constraint set is built by considering first a set of crisp constraints defining a set of feasible solutions $U \subset \mathscr{X}$, and a set of fuzzy constraints $\tilde{C}_i$, $i = 1, \dots m$. Let $\tilde{C}$ be the intersection of the $\tilde{C}_i$'s and U obtained using the intersection operation minimum.

Then we have to make sense of an optimal solution over the set of feasible solutions $\tilde{C}$ when the latter is fuzzy.

Following Bellman and Zadeh [1], one way is to build a fuzzy set from f and combine it with the fuzzy constraint set $\tilde{C}$ using again the intersection operation minimum. The simple way of constructing a fuzzy set is to consider the optimal value z^* of $f(x)$ in the support of the fuzzy constraint set $\{x : \mu_{\tilde{C}}(x) > 0\}$, and let the fuzzy optimizing set F be defined as $\mu_F(x) = \frac{f(x)}{z^*}$. Then we can define the fuzzy set of optimal solutions as $F \cap \tilde{C}$, an optimal solution x^* being the one with maximal membership grade in $F \cap \tilde{C}$, a path initiated by the late Tanaka and colleagues [15].

However this is just one way of defining the fuzzy optimizing set as there can be many increasing functions $\phi : \mathbb{R} \to \mathbb{R}$ with which we can define μ_F as $\frac{\phi(f(x))}{\phi(z^*)}$. So in order to properly define the fuzzy set $F \cap \tilde{C}$, we need to define a scaling function ϕ that makes the objective function and the membership function of the fuzzy constraint set commensurable. Clearly, the choice of function ϕ affects the choice of the optimal solution, which can be written x^*_ϕ. Zimmermann [19] found the way to choose a scaling function that makes sense, using a fuzzy expectation level on the objective function, hence giving up the idea of optimizing f. Namely, suppose the user can find thresholds $\underline{z}$ and $\overline{z}$ such that $f(x) \geq \overline{z}$ is judged sufficient for a solution x to be good enough, and completely insufficient if $f(x) \leq \underline{z}$. Then μ_F is often chosen as a linear increasing membership function such that $\mu_F(x) = 0$ if $f(x) \leq \underline{z}$, and 1 if $f(x) \geq \overline{z}$. In this case, if the fuzzy constraints are defined by linear membership functions, maximizing $\min(\mu_F(x), \mu_{\tilde{C}}(x))$ becomes a reasonable approach, whereby the objective function is handled as another fuzzy constraint. It readily extends to the case of several objective functions.

An alternative approach which dispenses with the choice of a scaling function was proposed by Verdegay [16], after Orlovsky [13], and sticks to the idea that objective functions and constraints, be they fuzzy, do not play the same role [16, 17]. Namely consider a feasibility threshold α and the crisp feasible set $C_\alpha = \{x : \mu_{\tilde{C}}(x) \geq \alpha\}$. It is then clear that we can get a standard optimization problem of maximizing $f(x)$ over C_α. The set of optimal solutions to this problem is defined by $S(\alpha) = \{x^* : f(x^*) = \sup_{x \in C_\alpha} f(x)\}$. Verdegay noticed that this is a form of parametric mathematical programming (MP) problem, in the sense that each choice of threshold α yields a different set $S(\alpha)$ of optimal solutions. The fuzzy set $\tilde{S}$ of optimal solutions to a fuzzy mathematical programming problem is then defined by Verdegay [16] by applying Zadeh's result for reconstructing a fuzzy set from its alpha-cuts:

$$\mu_{\tilde{S}}(x) = \sup\{\alpha : x \in S(\alpha)\}.$$

It is known since Orlovsky [13] (see also [6], pp. 102–103) that if $S = \cup_{\alpha>0} S(\alpha)$ then $\tilde{S} = S \cap \tilde{C}$. In contrast with the fuzzy set of solutions based on rescaling the objective function, this definition of a fuzzy solution to a fuzzy MP problem does not depend on any scaling function. It is up to the decision maker to choose the decision $\hat{x}$ that maximizes $\mu_{\tilde{S}}(x)$, or to choose some solution in the set $S(\alpha)$ for a suitable choice of α. This view is also very close to the parametric programming approach to fuzzy linear programming first proposed by the late Stefan Chanas [3] (see also [2]).

Verdegay [16] also showed that the two paradigms of fuzzy MP can be related if it is noticed that we can solve the fixed point problem $\psi(\alpha) = \sup_{x \in C_\alpha} \mu_F(x) = \alpha$, which has a solution under continuity conditions since ψ is decreasing with α. The fixed point solution α^* yields a set of solutions to the fuzzy MP $x^* \in S(\alpha^*)$ that do maximize $\min(\mu_F(x), \mu_{\tilde{C}}(x))$, i.e., correspond to the optimal solutions in the other paradigm.

The fuzzy solution of Verdegay is a fuzzy set constructed from crisp sets $S(\alpha)$ that are not necessarily nested. Indeed we do not have that if $\alpha > \beta$ then $S(\alpha) \subseteq S(\beta)$. This is because the optimal solutions in C_β may fail to lie in its subset C_α. So we can make two observations:

- The sets $S(\alpha)$'s are not nested. In fact, if $\alpha > \beta$, either $S(\alpha) \subseteq S(\beta)$ or $S(\alpha) \cap S(\beta) = \emptyset$, i.e., either $S(\alpha) = C_\alpha \cap S(\beta)$ or $S(\alpha) \cap S(\beta) = \emptyset$. Moreover, the $S(\alpha)$'s can be singletons if the optimal solutions are unique.
- The definition of a fuzzy optimal set in the sense of Verdegay is given by the mapping $\alpha \in (0, 1] \mapsto S(\alpha) \subset \mathscr{X}$, not originally a membership function $\mathscr{X} \to [0, 1]$.

This type of layered representation has been studied in the literature since Negoita and Ralescu [12] in the nested case, but the non-nested one has been more recently studied under the names of gradual sets [7], or RL-representations by Sanchez et al. [14] and by Martin and Azvine [9] in the so-called $X\mu$ approach. When $S(\alpha)$'s are singletons, the fuzzy solution corresponds to a *fuzzy element* whose definition is now recalled.

3 Fuzzy Elements and Gradual Numbers

Given a distributive lattice L with top 1 and bottom 0, and a set S, a fuzzy element $\tilde{s}$ is defined by a mapping $A_{\tilde{s}} : L^+ \to S$, where $L^+ = L \setminus \{0\}$ [7]. L is called a relevance scale, and $A_{\tilde{s}}$ an assignment function. The idea is that the choice of elements in S is parameterized by elements in L that are ordered in terms of relevance (e.g., quality, excellence, plausibility etc.): an element s is determined by $\alpha \in L^+$ in the sense that $A_{\tilde{s}}(\alpha) = s$; we may write $s_\alpha \in S$ for simplicity.

Note that an assignment function does not always determine a fuzzy set. However, given a fuzzy set F on S, with values in L, the inverse function μ_F^{-1} exists and is an assignment function if this membership function μ_F is bijective. So, some fuzzy sets can be viewed as fuzzy elements of the universe where the fuzzy set lies, formally. However, it is clear that beyond their possible mathematical identity, the two notions are distinct and will not be processed identically, e.g., fuzzy elements cannot be intersected like fuzzy sets do.

It is possible to view a fuzzy set as a fuzzy element of the power set $\wp(S)$ using alpha-cuts, namely, the mapping $\alpha \mapsto F_\alpha = \{s : \mu_F(s) \geq \alpha\}$ defines what can be called a gradual set, which is another description of a fuzzy set. However, there is no monotony constraint for a fuzzy element $\tilde{G}$ in $\wp(S)$, namely $\alpha \geq \alpha'$ does not imply $\tilde{G}(\alpha) \subseteq \tilde{G}(\alpha')$ like for fuzzy sets. Gradual sets are akin to so-called soft sets [11] (although the domain of a soft set is not supposed to be ordered, so that it is just the well-known notion of a relation), but they have been studied by other researchers, as pointed out in the previous section [9, 14].

A fuzzy set F can be viewed as a crisp set of fuzzy elements. Indeed, fuzzy elements are obtained by picking one element s_α in each cut F_α, namely $F = \{\tilde{s} : s_\alpha \in F_\alpha, \alpha \in L^+\}$, and we can show that the fuzzy set F is the set of its fuzzy elements via the reconstruction formula

$$\mu_F(s) = \max\{\alpha : s_\alpha = s\}.$$

A gradual real number $\tilde{r}$ is a fuzzy element of the reals, letting $S = \mathbb{R}$ and $L = [0, 1]$ [8]. Contrary to numbers, gradual real numbers are not totally ordered, and a partial order on gradual numbers can be naturally defined as follows: Let $\tilde{r}$ and $\tilde{s}$ be two gradual numbers. A gradual real number $\tilde{r}$ pointwisely dominates a gradual real number $\tilde{s}$ (written $\tilde{r} \geq \tilde{s}$) if and only if $\forall\alpha \in (0, 1]\, r_\alpha \geq s_\alpha$.

Given a fuzzy interval M, that is a fuzzy set of reals whose alpha-cuts are closed intervals $[\underline{m}_\alpha, \overline{m}_\alpha]$, we have that $M = \{\tilde{r} : \underline{m}_\alpha \leq r_\alpha \leq \overline{m}_\alpha, \alpha \in L^+\}$. As suggested in the introduction, the term *fuzzy number*, often used in place of fuzzy interval, may be misleading. Note that the difference between a gradual number and a fuzzy number in the usual acception is that the latter is a parametrized interval (depending on the choice of α) while a gradual number is a parameterized *number*, again controlled by α.

Consider two bijective functions

$$\mu_{M^-}(r) = \begin{cases} \mu_M(r) \text{ if } r \leq \underline{m}_1; \\ 1 \text{ otherwise.} \end{cases}$$

and

$$\mu_{M^+}(r) = \begin{cases} \mu_M(r) \text{ if } r \geq \overline{m}_1; \\ 1 \text{ otherwise.} \end{cases}$$

A fuzzy interval can be viewed as a crisp interval bounded by gradual numbers $\underline{\tilde{m}}$ and $\overline{\tilde{m}}$ defined by assignment functions $\mathscr{A}_{\underline{\tilde{m}}}(r) = \mu_{M^-}^{-1}(r)$ and $\mathscr{A}_{\overline{\tilde{m}}}(r) = \mu_{M^+}^{-1}(r)$.

Then we can view a fuzzy interval M as the interval $[\underline{\tilde{m}}, \overline{\tilde{m}}] = \{\tilde{r} : \underline{\tilde{m}} \leq \tilde{r} :\leq \overline{\tilde{m}}\}$. These gradual numbers have the shape of cumulative ($\underline{\tilde{m}}$) or survival ($\overline{\tilde{m}}$) probability functions, but can be considered as fuzzy thresholds offering a gradual transition from one half of the reals to the other half.

Operations on gradual numbers are defined in [8] as standard pointwise operations on assignment functions. For instance:

- The sum of two gradual numbers $\tilde{r}$ and $\tilde{s}$ is defined by summing their assignment functions. It is $\tilde{r} + \tilde{s}$ such that $\forall\alpha \in (0, 1]\, A_{\tilde{r}+\tilde{s}}(\alpha) = A_{\tilde{r}}(\alpha) + A_{\tilde{s}}(\alpha)$. For simplicity, we write $\forall\alpha \in (0, 1], (\tilde{r} + \tilde{s})(\alpha) = \tilde{r}(\alpha) + \tilde{s}(\alpha) = r_\alpha + s_\alpha$.
- The set of gradual numbers equipped with the addition operation forms a commutative group with identity $\tilde{0}$ such that $A_{\tilde{0}}(\alpha) = 0\,\forall\alpha \in (0, 1]$. The gradual number $\tilde{r}$ has the inverse $-\tilde{r}$ under the addition $\forall\alpha \in (0, 1]\, A_{-\tilde{r}}(\alpha) = -A_{\tilde{r}}(\alpha)$ and $\tilde{r} + (-(\tilde{r})) = \tilde{0}$.

- From the group structure, the subtraction operation is defined by $A_{\tilde{r}-\tilde{s}} = A_{\tilde{r}}+A_{-\tilde{s}} = A_{\tilde{r}} - A_{\tilde{s}}$.

Any algebraic operation on real numbers can be straightforwardly extended to gradual numbers. For details see [8]. Note that since gradual numbers are a fuzzy extension of numbers, not intervals, the algebraic structure of numbers is for the most part inherited by gradual numbers. This fact sheds useful light in the issue of confusing a fuzzy number understood as a fuzzy interval and a fuzzy number understood as a gradual number. For instance, a monotonic gradual number $\tilde{r}$ in the form of a cumulative distribution function (a form of fuzzy threshold used in fuzzy linear programming [19]) can be viewed as the gradual extension of a crisp number $r \in \mathbb{R}$, and the fuzzy extension of a semi-open interval $[r, +\infty)$. Under the first view it is natural to have that the difference $\mathscr{A}_{\tilde{r}-\tilde{s}}(\alpha) = r_\alpha - s_\alpha$, which is the gradual number extending $r - s$. In contrast, the result of the fuzzy subtraction $\tilde{r} \ominus \tilde{s}$ based on the extension principle, where we regard the monotonic gradual numbers as fuzzy intervals, is not obtained by computing $r_\alpha - s_\alpha$ as it is most of the time the whole real line. Namely, the interval subtraction

$$[r, +\infty) - [s, +\infty) = \{x - y : x \geq r, y \geq s\} = \mathbb{R}.$$

At this point it should be clear that the fuzzy solution to a fuzzy MP problem in the sense of Verdegay [16] is a very early example of a fuzzy element in (or a gradual subset of) the fuzzy constraint set, since we do have, by construction, that $S(\alpha) \subset C_\alpha, \forall \alpha \in (0, 1]$. This is just one example of gradual set or fuzzy element that can be found in the literature. In this case, it turns out that the specific features of such a fuzzy solution (any two sets in the family are nested or disjoint) allow us to view it as a fuzzy set from which we can recompute the gradual set via alpha-cutting. Other more recent examples of fuzzy elements and non-nested gradual sets are for instance the relative fuzzy cardinality of a fuzzy set [4], the fuzzy probability of a fuzzy event, the midpoint of a fuzzy interval [8], or the Haussdorf distance between fuzzy sets (see [7] for discussions).

4 Conclusion

This note has the only ambition to demonstrate the pioneering role of our colleague and friend J.L. Verdegay in the early times of fuzzy optimization, showing that he developed an alternative view to a fuzzy solution set, that he recalled in his recent retrospective position paper [18] and that would perhaps deserve to be studied further, as opposed to the popular max-min approach to fuzzy MP. It seems that recent research from Granada university focused on so-called SL-representations [4, 14]. The necessity to properly formalize fuzzy set related notions,such as fuzzy solutions to fuzzy optimization problems, led to define fuzzy elements in fuzzy sets and the like, where counterparts to alpha-cuts are no longer nested. Moreover, the connection to parametric programming also makes it clear that, just as a fuzzy solution is also a parametric solution driven by membership grades, a fuzzy set can be defined by letting a set depend on a parameter ranging on an ordered scale, and any scalar evaluation of this set defines a gradual element.

References

1. Bellman, R.E., Zadeh, L.A.: Decision making in a fuzzy environment. Manag. Sci. **17**, B141–B164 (1970)
2. Carlsson, C., Korhonen, P.: A parametric approach to fuzzy linear programming. Fuzzy Sets Syst. **20**, 17–30 (1986)
3. Chanas, S.: The use of parametric programming in fuzzy linear programming. Fuzzy Sets Syst. **11**, 243–251 (1983)
4. Delgado, M., Ruiz, M.D., Sánchez, D., Vila, M.A.: Fuzzy quantification: a state of the art. Fuzzy Sets Syst. **242**, 1–30 (2014)
5. Dubois, D., Fargier, H., Prade, H.: Possibility theory in constraint satisfaction problems: handling priority, preference and uncertainty. Appl. Intell. **6**, 287–309 (1996)
6. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and Applications, Mathematics in Science and Engineering Series, vol. 144. Academic Press, New York (1980)
7. Dubois, D., Prade, H.: Fuzzy elements in a fuzzy set. Soft Comput. **12**, 165–175 (2008)
8. Fortin, J., Dubois, D., Fargier, H.: Gradual numbers and their application to fuzzy interval analysis. IEEE Trans. Fuzzy Syst. **16**, 388–402 (2008)
9. Martin, T.P., Azvine, B.: The X-mu approach: Fuzzy quantities, fuzzy arithmetic and fuzzy association rules. In: Proceedings of IEEE Symposium on Foundations of Computational Intelligence (FOCI), pp. 24–29 (2013)
10. Meseguer P., Rossi F., Schiex T.: Soft Constraints, Chapter 9 in Foundations of Artificial Intelligence, vol. 2, pp. 281–328. Elsevier (2006)
11. Molodtsov, D.: Soft set theory—first results. Comput. Math. Appl. **37**(4/5), 19–31 (1999)
12. Negoita, C.V., Ralescu, D.A.: Applications of Fuzzy Sets to Systems Analysis. Birkhauser, Basel (1975)
13. Orlovsky, S.A.: Programming with fuzzy constraint sets. Kybernetes **6**, 197–201 (1977)
14. Sanchez, D., Delgado, M., Vila, M.A., Chamorro-Martinez, J.: On a non-nested level-based representation of fuzziness. Fuzzy Sets Syst. **192**, 159–175 (2012)
15. Tanaka, H., Okuda, T., Asai, K.: On fuzzy mathematical programming. J. Cybernet. **3**, 37–46 (1974)
16. Verdegay, J.L.: Fuzzy mathematical programming. In: Gupta, M.M., Sanchez, E. (eds.) Fuzzy Information and Decision Processes, North Holland, pp. 231–237 (1982)
17. Verdegay, J.L.: A dual approach to solve the fuzzy linear programming problem. Fuzzy Sets Syst. **14**(2), 131–141 (1984)
18. Verdegay, J.L.: Progress on fuzzy mathematical programming: a personal perspective. Fuzzy Sets Syst. **281**, 219–226 (2015)
19. Zimmermann, H.-J.: Description and optimization of fuzzy systems. Int. J. General Syst. **2**, 209–215 (1975)

Using Fuzzy Measures to Construct Multi-criteria Decision Functions

Ronald R. Yager

Abstract We are interested in the formulation of multi-criteria decision functions based on the use of a measure over the space of criteria. Specifically the relationship between the criteria is expressed using a fuzzy measure. We then use the Choquet integral to construct decision functions based on the measure. We look at a number of different decision functions generated from specific classes of measures.

1 Introduction

One approach to multi-criteria decision-making is to construct decision function by aggregating an alternative's satisfaction to the individual criteria and then selecting the alternative with the largest aggregated value of the individual criteria [1]. Our focus here is on the formulation of multi-criteria decision functions where our aggregation method is based on the use of a fuzzy measure (monotonic set measure) and the Choquet integral [2–5]. In this framework the measure is used to convey information about criteria importance's and relationships between the constituent criteria. We first describe the general approach to the formulation of decision functions using this framework. We look at the types of aggregation functions that are generated from various classes of measures. We show to how to aggregate the underlying measures to enable the modeling of more complex relationships between the criteria from simple relationships.

R.R. Yager (✉)
Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, USA
e-mail: yager@panix.com

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_14

2 Measure Based Approach to Multi-criteria Decision Making

Assume we have a collection $\mathbf{C} = \{C_1, \cdots, C_n\}$ of criteria of interest in a decision problem. Let X be a set of alternatives from among which we must select the one that best satisfies the criteria. Here for each alternative x we let $C_i(x) \in [0, 1]$ indicate the degree to which criteria C_i is satisfied by alternative x. In order to select the alternative that best satisfies the collection $\mathbf{C}$ we must provide some function F that indicates the degree to which each x satisfies the collection of criteria. We shall denote this as $F(x) = Agg(C_1(x), C_2(x), \cdots, C_n(x))$.

A very general formulation for Agg can be obtained with the aid of a fuzzy measure on the space of criteria and the use of an appropriate integral [3]. In this approach the fuzzy measure is used to express the structural relationship between the criteria. A fuzzy measure on the space $\mathbf{C}$ of criteria is a mapping μ: $2^{\mathbf{C}} \rightarrow [0, 1]$ such that

$$\mu(\mathbf{C}) = 1, 2.\mu(\varnothing) = 0 \text{ and } 3.\mu(A) \geq \mu(B) \text{ if } A \supseteq B \tag{1}$$

An interpretation of μ in this environment of multi-criteria decision-making is that for any subset A in C, $\mu(A)$ is the importance associated with the subset of criteria in A.

As we indicated we shall use the fuzzy measure to guide the construction of the aggregation function $F(x) = Agg(C_1(x), \cdots, C_n(x))$. One general approach for obtaining a decision function F that makes use of this fuzzy measure on the collection of criteria is the Choquet integral, $F(x) = Ch_\mu(C_1(x), C_2(x), \cdots, C_n(x))$ [2–5]. In anticipation of introducing the Choquet integral we provide some formalism. For a given alternative x we let id(j) be the index of the jth most satisfied criteria. Thus $C_{id(j)}(x)$ is the degree of satisfaction of the jth most satisfied criteria. We now let $H_j = \{C_{id(k)} | k = 1 \text{ to } j\}$, it is the collection of the j most satisfied criteria. Here we see $H_n = \mathbf{C}$. We shall let $H_0 = \varnothing$ by convention. We see here that $H_k \subseteq H_j$ for $j \geq k$. We further see from the monotonicity of μ that $\mu(H_j)$ is a monotonically non-decreasing function of j, $\mu(H_i) \geq \mu(H_k)$ if $i \geq k$. Using the $\mu(H_j)$ we can obtain a collection of weights, $w_j = \mu(H_j) - \mu(H_{j-1})$, for j = 1 to n. It is easily to show that $w_j \geq 0$ for all j and $\sum_{j=1}^{n} w_j = 1$. Using this we get $F(x) = Ch_\mu(C_1(x), C_2(x), \cdots, C_n(x)) = \sum_{j=1}^{n} w_j C_{id(j)}(x)$.

It is easy to show that this type of aggregation function is a mean. In particular it is known that F(x) is bounded, $Min_i[C_i(x)] \leq F(x) \leq Max_i[C_i(x)]$, and monotonic, if $C_i(x) \geq C_i(y)$ for all i then $F(x) \geq F(y)$. Another property we can show is that if $\mu_1(A) \geq \mu_2(A)$ for all A then for x, $F_1(x) \geq F_2(x)$.

3 Combining Measures

We shall make some general observations about fuzzy measures. First we recall the definition of an aggregation function [5]. A mapping Agg: $I^n \rightarrow I$ is called an aggregation function if it satisfies the three conditions. (1) $Agg(0, 0, \cdots, 0) = 0$, (2) $Agg(1, \cdots, 1) = 1$ and (3) $Agg(a_1, \cdots, a_n) \geq Agg(b_1, \cdots, b_n)$ if $a_i \geq b_i$ for all i.

We note that there are many aggregation functions [5], among the more notable of these are the Max, the Min and the average. We also note that product is an aggregation function. All mean operators are aggregation functions as are t-norms and t-conorms [5].

In the following we note a fundamental theorem of fuzzy measures [6].

Theorem: Assume $\mu_1, \ldots, \mu_q$ are a collection of fuzzy measures on the space Z. If Agg is an aggregation function then the set function μ defined such that $\mu(A) = Agg(\mu_1(A), \cdots, \mu_n(A))$ for all subsets A of Z is itself a fuzzy measure.

We shall refer to this as FTAM, the Fundamental Theorem on Aggregation of Measures. The FTAM provides a very general approach to constructing set measures from other measures. We shall use the notation $\mu = Agg(\mu_1, \ldots, \mu_n)$ to indicate that μ is defined so that $\mu(A) = Agg(\mu(A), \ldots, \mu(A))$ for all subsets A.

Since product is an aggregation function then $\mu = \mu_1\mu_2, \ldots, \mu_n$ is a measure. Also we note that $\mu = Max(\mu_1, \ldots, \mu_n)$ is a measure as well as $Min(\mu_1, \ldots, \mu_n)$. Furthermore if w_j for j = 1 to j are such that $w_j \in [0, 1]$ and $\Sigma\, w_j = 1$ then $\mu = \sum_{j=1}^{n} w_j\mu_j$ is a measure. Here $\mu(A) = \sum_{j=1}^{n} w_j\mu_j(A)$ for all subsets A of Z.

We now recall that the Choquet integral generates an aggregation function. Thus this can provide a methodology for constructing new measures. Let $\mathbf{C} = \{C_1, \cdots, C_q\}$ be a collection of criteria. Let $R = \{\mu_1, \ldots, \mu_r\}$ be a collection of measures on the space **C** of criteria. Let m be a measure on the space R. We now use this to form a compound measure μ on **C** defined so that for any subset A of the criteria **C** we have

$$\mu(A) = Choq_m(\mu_1(A), \mu_2(A), \cdots, \mu_r(A))$$

Here μ(A) is the Choquet integral with respect to m with arguments $\mu_j(A)$. We now show how this measure μ can be used to determine the overall satisfaction of alternative x, F(x).

Assume $C_i(x) = a_1$, is the satisfaction of criteria C_i by alternative x. Without loss of generality we shall assume the indexing has done so that $a_i \geq a_k$ if $i < k$. In this case $H_j = \{C_1, C_2, \cdots, C_j\}$.

Using the Choquet integral we have $F(x) = F(a_1, \cdots, a_n) = \sum_{j=1}^{n} w_j a_j$ where $w_j = \mu(H_j) - \mu(H_{j-1})$. However here $\mu(H_j) = Choq_m(\mu_1(H_j), \mu_2(H_{j)}, \cdots, \mu_r(H_{j)}.$

4 Basic Weighted Average Aggregation

In the following we shall look at the types of multi-criteria decision functions we get using the Choquet integral under some notable examples of fuzzy measures. The most basic example is the additive measure. For this measure we define μ as follows. For each C_i we associate a value α_i such $\alpha_i \in [0, 1]$ and $\sum_{i=1}^{q} \alpha_i = 1$. Here for any subset $A \subseteq \mathbf{C}$ we have $\mu(A) = \sum_{k, C_k \in A} \alpha_k$. Here if $A(C_k)$ is the membership grade of C_k in A, $A(C_k) = 1$ if $C_k \in A$ and $(C_k) = 0$ if $C_k \notin A$ then $\mu(A) = \sum_{k=1}^{q} \alpha_k A(C_k)$

Let us now obtain the Choquet integral in this case. If $C_i(x)$ is the satisfaction of C_i by x then $F(x) = \text{Choq}_\mu[C_1(x), \cdots, C_q(x)]$. Let id(j) be the index of jth largest of the $C_i(x)$ using this we have $H_j = \{C_{id(k)}/k = 1 \text{ to } j\}$. In this case $F(x) = \sum_{j=1}^{q} (\mu(H_j) - \mu(H_{j-1}))C_{id(j)}(x)$. Since $\mu(H_j) = \sum_{k=1}^{j} \alpha_{id(k)}$ and $\mu(H_j) = \sum_{k=1}^{j-1} \alpha_{id(k)}$ then $\mu(H_j) - \mu(H_{j-1}) = \alpha_{id(j)}$ and hence $F(x) = \sum_{j=1}^{q} \alpha_{id\,(j)} C_{id\,(j)}(x) = \sum_{i=1}^{q} \alpha_i C_i(x)$. It is the simple weighted average of the satisfactions where the weight associated with criteria C_i is α_i. Here we have the notable feature that for a given criteria, C_i, no matter what position it appears in the ordering id its associated weight is always α_i. Thus in this case it appears justifiable to refer to α_I as the importance associated with C_i.

5 Cardinality Based Measures and OWA Aggregation

An important class of measures studied by Yager [7] are the cardinality based measures. A fuzzy measure μ is called a cardinality based measure if $\mu(A) = V_{|A|}$. Here the measure of a subset just depends upon the number of elements in it. It is understood here if $A = \varnothing$, $\mu(\varnothing) = 0$ and hence $V_0 = 0$ and since $\mu(C) = 1$ we have $V_n = 1$. Thus we see that a cardinality based measure is defined by a set of values $0 = V_0 \leq V_1 \leq V_2 \leq \cdots \leq V_n = 1$ such that $\mu(A) = V_{|A|}$.

Let us see the Choquet integral in the case of a cardinality-based measure. Since

$$F_\mu(x) = \sum_{J=1}^{n} (\mu(H_j) - \mu(H_j - 1))C_{id(j)}(x)$$

where H_j is the set of the j criteria with the largest satisfaction to x. We note that since the cardinality of H_j is j thus $F_\mu(x) = \sum_{J=1}^{n} (V_j - V_{j-1})C_{id(j)}(x)$. If we denote $V_j - V_{j-1} = w_j$ we have $F_\mu(x) = \sum_{J=1}^{n} w_j C_{id(j)}(x)$ where each $w_j \geq 0$ and $\sum_{J=1}^{n} w_j = 1$. We see here that this is the OWA aggregation operator introduced by Yager [8].

A number of important examples of this case are (1) μ^* where $V_0 = 0$ and $V_j = 1$ for $j > 1$, (2) μ_* where $V_n = 1$ and $V_j = 0$ for $j < n$ and (3) μ_S where $V_j = j/n$ for all j. We easily see that

(1) $\mu^* \Rightarrow F_{\mu^*}(x) = \text{Max}_i[C_i(x)]$
(2) $\mu_* \Rightarrow F_{\mu_*}(x) = \text{Min}_i[C_i(x)]$
(3) $\mu_S \Rightarrow F_{\mu S}(x) = \frac{1}{n}\sum_{i=1}^{n} C_i(x)$

In [9] Yager suggested we can obtain the parameters for this the type of cardinality based measure using a function called a weight generating function g: [0, 1] → [0, 1] having the properties: (1) $g(0) = 0$, (2) $g(1) = 1$ and (3) $g(x) \geq g(y)$ if $x > y$ (monotonicity). Using this weight generating function we can obtain

$$V_j = g\left(\frac{j}{n}\right) \text{ and } w_j = g\left(\frac{j}{n}\right) - g\left(\frac{j-1}{n}\right).$$

We note a special case of g is linear, $g(x) = x$. Here we get $V_j = j/n$ and $w_j = 1/n$.

In [9] Yager discussed various semantics that can be associated with g. One particularly notable semantics is where g is a quantifier indicating the proportion of criteria that must be satisfied.

Earlier we showed that if μ_1 and μ_2 are two measures such that $\mu_1(A) \geq \mu_2(A)$ for all A then if $F_1(x)$ and $F_2(x)$ are the respectively Choquet integrals obtained using these measures then $F_1(x) \geq F_2(x)$ for all x. Let us look at the implication of this for the case of cardinality-based measures and the related OWA operator. If μ_1 and μ_2 are two cardinality based measures such that V_{1k} and V_{2k} are their respected parameters then $\mu_1(A) = V_{1|A|}$ and $\mu_2(A) = V_{2|A|}$ and if $V_{1k} \geq V_{2k}$ for $k = 1$ to n we have $F_1(x) \geq F_2(x)$ for all x.

Consider now an OWA operator defined in terms of a collection of weights $w_1, \ldots, w_n$. From the preceding we see that this is equal to a cardinality-based formulation in which $V_j = \sum_{k=1}^{j} w_k$. From this we can conclude the following. Assume w_{ik} and w_{2k} are two collections of OWA weights. Let $OWA_1(C_1(x), \cdots, C_n(x))$ and $OWA_2(C_1(x), \cdots, C_n(x))$ be the OWA aggregations under these respective weights. Then we see that if for all j we have that $\sum_{k=1}^{j} w_{1k} \geq \sum_{k=1}^{j} w_{2k}$ then $OWA_1(C_1(x), \cdots, C_n(x)) \geq OWA_2(C_1(x), \cdots, C_n(x))$. Furthermore if g_1 and g_2 are two weight generating functions such that $g_1(y) \geq g_2(y)$ for all $y \in [0, 1]$ then the aggregation obtained using g_1 will always be at least as large as that obtained using g_2.

A related class of measures can be obtained using a function g and a set of weights, α_j, associated with each C_j such that $\alpha_j \in [0, 1]$ and $\Sigma_j \alpha_j = 1$. Here α_j is seen as some kind of importance associated with criterion j. Using this information we define $\mu(A) = g\left(\sum_{C_i \in A} \alpha_i\right)$. We see in this case $\mu(H_j) = g\left(\sum_{i, C_i \in H_j} \alpha_i\right)$ and hence

$$w_j = g\left(\sum_{i\in H_j} \alpha_i\right) - g\left(\sum_{i\in H_{j-1}} \alpha_i\right).$$

If we let id(j) denote the index of the jth most satisfied criteria then $\mu(H_j) = g = \left(\sum_{k=1\,\text{to}\,j} \alpha_{id(k)}\right)$ where $\alpha_{id(k)}$ is the importance weight associated for kth most satisfied criteria. Using this notation we see that

$$F(x) = \sum_{j=1}^{n}\left(g\left(\sum_{k=1}^{j} \alpha_{id(k)}\right) - g\left(\sum_{k=1}^{j-1} \alpha_{id(k)}\right)\right) C_{id(k)}(x).$$

In the special case when g is linear, g(x) = x, then we see that $F(x) = \sum_{k=1}^{n} \alpha_{id(k)} C_{id(k)}(x) = \sum_{i=1}^{n} \alpha_i C_i(x)$. It is simply the importance weighted average.

6 Prioritized Multi-criteria Aggregation

An important type of relationship between criteria is illustrated by the following example. Consider we are choosing a bicycle for a child and we have two criteria of interest, safety and price. Assume the decision maker's preference is that the safety is of utmost importance. In particular, he is not willing to let high satisfaction to the criteria of price compensate for poor satisfaction to the criteria of safety. Here we say that safety has a ***priority*** over cost and denote this Safety > Cost.

In [10] we suggested a formulation for a fuzzy measure that can be used to implement a priority relationship between the criteria. Assume $\mathbf{C} = \{C_1, \cdots, C_n\}$ are prioritized so that $C_1 > C_2 > \cdots > C_n$. As noted above our basic idea of prioritization is that lack of satisfaction to higher priority criteria is not easily compensated by satisfaction to lesser priority criteria. In the following we introduce a measure to implement this type of imperative. We first define $L_j = \{C_k \mid k = 1 \text{ to } j\}$ for j = 1 to n and $L_0 = \varnothing$. We now associate with each subset L_j a value $V_j = j/n$. Using this we define the measure μ such that $\mu(A) = \text{Max}_{j=1\,\text{to}\,n}\left[V_j G_j(A)\right]$ where $G_j(A) = 1$ if $L_j \subseteq A$ and $G_j(A) = 0$ if $L_j \not\subset A$. We see that $\mu(A) = j/n$ where L_j is the largest L_j that is contained in A. We easily see that $\mu(\varnothing) = 0$, $\mu(\mathbf{C}) = 1$ and $\mu(A) \geq \mu(B)$ if $A \supseteq B$. Thus μ is a fuzzy measure.

Let us look at μ for some subsets of **C**. Consider the case of singleton sets $\mu(\{C_k\})$. We see that $\mu(\{C_1\}) = 1/q$ while $\mu(\{C_k\}) = 0$ for $k \neq 1$. Thus only the singleton set consisting of C_1, the highest priority element, has a non-zero measure. In the case of subsets A consisting of two criteria:

$$\mu(\{A\}) = 2/q \text{ if } A \cap \{C_1,\ C_2\}) = \{C_1,\ C_2\}$$

$$\mu(\{A\}) = 1/q \text{ if } A \cap \{C_1, C_2) = \{C_1\}$$

$$\mu(A) = 0 \text{ if } A \cap \{C_1) = \varnothing.$$

Additionally we see

$$\mu(\{C_1, C_2\}) = 2/q$$

$$\mu\{C_1, C_k)\} = 1/q \text{ for } C_k \neq C_2$$

$$\mu\{C_i, C_k) = \varnothing \quad \text{if neither i or k is 1}$$

We observe that for any subset A such that $C_1 \notin A$ then $G_j(A) = 0$ for all j and $\mu(A) = 0$, for any other subset A, $\mu(A)$ is equal to V_j where L_j is the maximum L_k contained in A.

We now shall investigate the use of the Choquet integral to obtain an aggregation function using this prioritization type measure. Here $F(x) = Agg(C_1(x), \ldots, C_n(x)) = \sum_{j=1}^{n} (\mu(H_j) - \mu(H_{j-1}))C_{id(j)}(x)$ where $C_{id(j)}$ is the jth most satisfied criteria and $H_j = \{C_{id(k)} \mid k = 1 \text{ to } j\}$, the collection of the j criteria with the largest satisfactions. Letting $w_j = \mu(H_j) - \mu(H_{j-1})$ we have $F(x) = \sum_{j=1}^{n} w_j C_{id(j)}(x)$. We note that since μ is monotonic then F(x) is monotonic in the $C_i(x)$ and also we have $Min_i[C_i(x)] \leq F(x) \leq Max_i[C_i(x)]$.

Consider the case where the highest priority criteria, C_1, is the least satisfied criteria. Here we have that $C_1 \notin H_j$ for $j = 1$ to $n - 1$ and only $C_1 \notin H_n$. In this case $\mu(H_j) = 0$ for $j = 1$ to $n - 1$ and $\mu(H_n) = 1$. Here then $Agg(C_1(x), \cdots, C_n(x)) = C_1(1)$. Since $C_1(x) = Min_i[C_i(x)])$ then here we have $Agg(C_1(x), \cdots, C_n(x) = Min_j[C_j(x)]$. It is the smallest value and there is no compensation by any other criteria.

Consider now the more general case where $C_1(x)$ is the pth largest of the satisfactions. Here $C_1 \notin H_j$ for $j = 1$ to $p - 1$ and thus $\mu(H_j) = 0$ for $j = 1$ to $p - 1$. From this we conclude $F(x) = \sum_{j=1}^{n} w_j C_{id(j)}(x) = \sum_{j=P}^{n} w_j C_{id(j)}(x)$. Furthermore since for $j = p + 1$ to n we have that $C_{id(j)}(x) \leq C_1(x)$ combining this with the fact that $\sum_{j=1}^{n} w_j = 1$ then we have $F(x) = \sum_{j=1}^{n} w_i C_{id(j)}(x) \leq \sum_{j=P}^{n} w_j C_{id(j)}(x) \leq C_1(x)$. We see that it is always the case in this priority aggregation that $Agg(C_1(x), \cdots, C_n(x)) \leq C_1(x)$.

Consider the case where $ind(j) = j$, the satisfactions are ordered the same as the priority. In this case $H_j = \{C_1, .., C_j\} = L_j$ and therefore $w_j = \mu(H_j) - \mu(H_j - 1) = \mu(L_j) - \mu(L_j - 1) = \frac{1}{n}$. Thus here we have $F(x) = \frac{1}{n}\sum_{j=1}^{n} C_j(x)$. It is the average of all the criteria satisfactions.

A slightly more general formulation of this prioritized aggregation can be had. Again assume $C_1 > C_2 > \cdots > C_n$ and let $L_j = \{C_k \mid k = 1 \text{ to } j\}$ and $L_0 = \varnothing$. Here we associate with each L_j a value $\lambda_j \geq 0$ such that $\lambda_i \geq \lambda_k$ for $i > k$ and $\lambda_n = 1$. We now define our measure μ such that $\mu(A) = Max_{j_{j=1 \text{ to } n}} [\lambda_j G_j(A)]$ where

$$G_j(A) = 1 \quad \text{if } L_j \subseteq A$$
$$G_j(A) = 0 \quad \text{if } L_j \not\subseteq A$$

So here we have a priority allowing different weights. In this case using the Choquet integral we again get $F(x) = \sum_{j=1}^{n} (\mu(H_j) - \mu(H_{j-1}))C_{id(j)}(x)$ with $H_j = \{C_{id(k)}/k = 1 \text{ to } j)$. An interesting special case is where $\lambda_j = 0$ for j = 1 to n − 1 and $\lambda_n = 1$. In this case we get $F(x) = \text{Min}_i[C_i(x)]$.

We now briefly consider a situation closely related to a prioritization of criteria. Assume C_1 and C_2 are two criteria such that for C_1 to be of any use we must satisfy criteria C_2. Here we say criterion C_1 requires criterion C_2. We can represent this using a measure μ by specifying for that any subset A if $A \cap \{C_2\} = \varnothing$ then $\mu(A \cup \{C_1\}) = \mu(A)$. Consider now the case where $C_1(x) > C_2(x)$. Here then if id (j) = 1 then id(k) = 2 for k > j. Here we see that $C_2 \not\subset H_j$ and $C_2 \not\subset H_{j-1}$ where $C_1 \subseteq H_j$ and $C_j \not\subset H_{j-1}$. Furthermore the weight associated with C_1, $C_{id(j)}$, $w_j = \mu(H_j) - \mu(H_{j-1})$. Since $\{C_2\} \cap H_{j-1} = \varnothing$ then $\mu(H_j) = \mu(H_{j-1} \cup \{C_1\}) = (H_{j-1})$ and hence $w_j = 0$. Thus we see for the case $C_1(x) > C_2(x)$ the weight associated with C_1 is zero, it makes no contribution.

7 Multi-criteria Aggregation Based on Quasi-additive Measures

Another class of measures useful for modeling multi-criteria decision function are the "quasi-additive" measures. Here we associate with the space **C** of criteria a collection $S_1, .., S_r$ of subsets. We note that these subsets need not be disjoint or that their union covers **C** We further associate with each S_j a value $\alpha_j \in [0, 1]$ such that $\sum_{j=1}^{r} \alpha_j = 1$. Using this we define the measure μ^S on **C** such that $\mu^S(A) = \sum_{j=1}^{r} R_j(A)\alpha_j$ where

$$R_j(A) = 1 \text{ if } A \cap S_j \neq \varnothing \text{ and } R_j(A) = 0 \text{ if } A \cap S_j = \varnothing$$

Here we are giving an importance weight of α_j to getting satisfaction to *any* criteria in S_j. This called a plausibility measure.

Let us consider the aggregation of criteria satisfactions under μ^S using the Choquet integral. Again we have

$$F(x) = \sum_{j=1}^{r} (\mu^S(H_j) - \mu^S(H_j - 1))C_{id(j)}(x) = \sum_{j=1}^{r} w_j C_{id(j)}(x),$$

Consider now the special case where $S_j = \{C_j\}$ for j = 1 to n. Here with $H_j = \{C_{id(k)}/k = 1 \text{ to } j\}$ we have $\mu^S(H_j) = \sum_{k=1}^{j} \alpha_{id(j)}$ and with $H_{j-1} = \{C_{id(k)}/ k = 1 \text{ to } j-1\}$ we have $\mu^S(H_{j-1}) = \sum_{k=1}^{j-1} \alpha_{id(j)}$. Here $w_j = \alpha_{id(j)}$ and we get

$F(x) = \sum_{j=1}^{n} \alpha_{id_{(j)}} C_{id(j)}(x) = \sum_{i=1}^{n} \alpha_i C_i(x)$. Thus in this case we get the simple importance weighted criteria satisfaction as a special case.

Another special case is where r = 1 and $\alpha_1 = 1$. Here we just have one subset S_1. Here we see $\mu^S(H_j) = 1$ if $S_1 \cap H_j \neq \emptyset$ and $\mu^S(H_j) = 0$ if $S_1 \cap H_j = \emptyset$. We see that in this case $\mu^S(H_j) = 1$ the first time we get an element from S_1 in H_j. Thus here $F(x) = \underset{j \in S_1}{\mathrm{Max}}\left[C_j(x)\right]$. Thus it is the value of the maximally satisfied criteria in S_1.

Another special case is where $S_1 = C$ and S_2 some arbitrary subset. Here we can show that

$$F(x) = \alpha_1 \mathrm{Max}_i[C_i(x)] + (1 - \alpha_1 \underset{j \in S_2}{\mathrm{Max}}[C_i(x)]$$

A related measure can be obtained if we define $R_j(A) = 1$ if $S_j \subseteq A$ and $R_j(A) = 0$ if $S_j \not\subset A$ and we define μ_S such that $\mu_S(A) = \sum_{j=1}^{r} R_j(A)\alpha_j$. Here we are giving an importance weigh α_j for satisfying *all* the criteria in S_j.

We can show in this case that if $S_j = \{C_j\}$ for j = 1 to n then this also reduces $F(x) = \sum_{j=1}^{n} \alpha_i C_i(x)$. We also can show that in the case where $\alpha_1 = 1$ then $F(x) = \mathrm{Min}_{C_j \in S_1} C_j(x)]$.

References

1. Bellman, R.E., Zadeh, L.A.: Decision-making in a fuzzy environment. Manag. Sci. **17**(4), 141–164 (1970)
2. Mesiar, R., Mesiarová, A.: Fuzzy integrals—what are they ? Int. J. Intell. Syst. **23**, 199–212 (2008)
3. Klement, E.P., Mesiar, R., Pap, E.: A universal integral as common frame for Choquet and Sugeno. IEEE Trans. Fuzzy Syst. **18**, 178–187 (2010)
4. Grabisch, M., Murofushi, T., Sugeno, M.: Fuzzy Measures and Integrals. Physica-Verlag, Heidelberg (2000)
5. Beliakov, G., Pradera, A., Calvo, T.: Aggregation Functions: A Guide for Practitioners. Springer, Heidelberg (2007)
6. Yager, R.R.: A measure based approach to the fusion of possibilistic and probabilistic uncertainty. Fuzzy Optim. Decis. Making **10**, 91–113 (2011)
7. Yager, R.R.: On the cardinality index and attitudinal character of fuzzy measures. Int. J. Gen. Syst. **31**, 303–329 (2002)
8. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Trans. Syst. Man Cybern. **18**, 183–190 (1988)
9. Yager, R.R.: Quantifier guided aggregation using OWA operators. Int. J. Intell. Syst. **11**, 49–73 (1996)
10. Yager, R.R.: On prioritized multiple criteria aggregation. IEEE Trans. Syst. Man Cybern. B **42**, 1297–1305 (2012)

A Modal Account of Preference in a Fuzzy Setting

Francesc Esteva, Lluís Godo and Amanda Vidal

Abstract In this paper we first consider the problem of extending fuzzy (weak and strict) preference relations, represented by fuzzy preorders on a set to a fuzzy preferences on subsets, and we characterise different possibilities. Based on their properties, we then semantically define and axiomatize several two-tiered graded modal logics to reason about the corresponding different notions of fuzzy preferences.

1 Introduction

Reasoning about preferences is a hot topic in Artificial Intelligence since many years, see for instance [5, 17, 18]. Two main approaches for representing and handling preferences have been developed: the relational and the logic-based approaches.

This paper is our humble contribution to the tribute, in the occasion of his 65th birthday, to José Luis "Curro" Verdegay. Excellent researcher and better person, he has been one of the pioneers of fuzzy logic in Spain and founder and driving force of the research group on Decision Making and Optimization at the University of Granada. Our contribution is devoted to logic and fuzzy preferences, a topic that, although it is not central on the research of Curro, is ubiquitous in fuzzy decision making models and we hope it may be of his interest. Along many years, we have jointly participated in many events around the world with Curro and with our friends from Granada, we have learnt a lot from his research ideas and organizational competences, but more importantly, we have enjoyed his friendship and shared many unforgettable moments. Thanks for all Curro, and congratulations for this well-deserved homage!

F. Esteva · L. Godo (✉)
IIIA - CSIC, 08193 Bellaterra, Spain
e-mail: godo@iiia.csic.es

F. Esteva
e-mail: esteva@iiia.csic.es

A. Vidal
ICS, Czech Academy of Sciences, Prague, Czech Republic
e-mail: amanda@cs.cas.cz

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_15

In classical preference relations, every preorder R (and more in general every reflexive relation) can be regarded as a preference relation by assuming that $(a, b) \in R$ means that *a is preferred or indifferent to b*. From R we can define three disjoint relations:

- the *strict preference* $P = R \cap R^d$,
- the *indifference relation* $I = R \cap R^t$, and
- the *incomparability relation* $J = R^c \cap R^d$.

where $R^d = \{(a, b) \in R : (b, a) \notin R\}$, $R^c = \{(a, b) \in R : (b, a) \in R\}$ and $R^t = \{(a, b) : (b, a) \in R\}$. It is clear that P is a strict order (irreflexive, antisymmetric and transitive), I is an equivalence relation (reflexive, symmetric and transitive) and J is irreflexive and symmetric. The triple (P, I, J) is called a *preference structure*, where the initial weak preference relation can be recovered as $R = P \cup I$.

In the fuzzy setting, preference relations can be attached degrees (usually belonging to the unit interval $[0, 1]$) of fulfilment or strength, so they become *fuzzy relations*. A weak fuzzy preference relation on a set X will be now a fuzzy preorder $R : X \times X \to [0, 1]$, where $R(a, b)$ is interpreted as the degree in which b is at least as preferred as a. Given a t-norm $\odot$, a fuzzy $\odot$-preorder satisfies reflexivity ($R(a, a) = 1$ for each $a \in X$) and $\odot$-transitivity ($R(a, b) \odot R(b, c) \leq R(a, c)$ for each $a, b, c \in X$). The most influential reference is the book by Fodor and Roubens [6], that was followed by many other works like, for example [7–11]. The problem in this setting is how to define the corresponding strict preference, indifference and incomparability relations from the initial fuzzy preorder. Many questions arise since it is possible to generalise the classical case in many different ways. In particular, several works have paid attention to how suitably interrelate a weak preference (a fuzzy preorder) with its associated indifference relation (a indistinguishability relation) and strict preference (a strict fuzzy order). In this sense, relevant publications are, among others, Bodenhofer's papers [2–4]. There, the author studies $\odot$-E fuzzy preorders related to a t-norm $\odot$ and an indistinguishability, or *fuzzy equivalence*, relation E (reflexive, symmetric and $\odot$-transitive), as well as their strict associated fuzzy orders in a general context, which is also applies to the context of preference modelling. Indeed, given a t-norm $\odot$ and an indistinguishability relation E, a $\odot$-E fuzzy preorder is defined as a fuzzy relation $R : X \times X \to [0, 1]$ satisfying: *E-reflexivity*: $R(x, y) \geq E(x, y)$, $\odot$-*E-antisymmetry*: $R(x, y) \odot R(y, x) \leq E(x, y)$, $\odot$-*transitivity*: $R(x, y) \odot R(y, z)) \leq R(x, z)$. Bodenhofer also studies how to extend such a $\odot$-E fuzzy preorder to the set $\mathcal{F}(\mathcal{X})$ of fuzzy subsets of a universe X, as well as the associated indistinguishability relation and the strict fuzzy order, and discusses different possible definitions. In such a setting, he considers both the cases of crisp and fuzzy preorders, but he does not consider the particular case we will study in this paper, namely the interaction of a fuzzy preferences over crisp subsets of X.

The basic assumption in logical approaches is that preferences have structural properties that can be suitably described in a fomalized language.This is the main goal of the so-called *preference logics*, see e.g. [17]. The first logical systems to reason about preferences go back to Halldén [20] and to von Wright [16, 22, 23]. More recently van Benthem et al. in [1] have presented a modal logic-based formalization of preferences. In that paper the authors first define a basic modal logic with two unary modal operators $\Diamond^{\leq}$ and $\Diamond^{<}$, together with the universal and existential modalities, A and E respectively, and axiomatize it. Using these primitive modalities, they consider several (definable) binary modalities to capture different notions of preference relations on classical propositions, and show completeness with respect to the intended preference semantics. Finally they discuss their systems in relation with von Wright axioms for *ceteris paribus* preferences [22]. On the other hand, with the motivation of formalising a comparative notion of likelihood, Halpern studies in [15] different ways to extend preorders on a set X to preorders on subsets of X and their associated strict orders. He studies their properties and relations among them, and he also provides an axiomatic system for a *logic of relative likelihood*, that is proved to be complete with respect to what he calls *preferential structures*, i.e. Kripke models with preorders as accessibility relations.

In this paper we begin by studying in Sect. 2 different forms to define fuzzy relations on the set $\mathcal{P}(\mathrm{W})$ of subsets of W, from a fuzzy preorder on W, in a similar way to the one followed in [1, 15] for classical preorders, and in [2, 3] for fuzzy preorders. In Sect. 3 we characterize them and discuss which are the most appropriate from the point of view of preference modelling, while in Sect. 4 we deal with the problem of defining a fuzzy strict order in a set associated to a given fuzzy preorder, and how to lift them to susbsets. Finally, in Sect. 5, and based on the previous results, we semantically define and axiomatize several two-tiered graded modal logics to reason about different notions of preferences.

This paper is a proper extended version of the conference paper [13].

2 Extending a Fuzzy Preorder on a Set W to a Fuzzy Relation on Subsets of W

2.1 Precedents in the Classical Case

In the classical logic setting, van Benthem et al. define in [1] *preference models* as triples $\mathcal{M} = (\mathrm{W}, \preceq, \mathcal{V})$ where W is a set of states or worlds, $\preceq$ is a preorder (reflexive and transitive) relation on W, and $\mathcal{V}$ is a standard propositional evaluation, that is, a mapping assigning to every propositional variable p a subset $\mathcal{V}(p) \subseteq \mathrm{W}$ of states where p is true. $\mathcal{V}$ can be extended to any propositional formula φ by using the classical Boolean definitions. For simplicity, we will also denote $\mathcal{V}(\varphi)$ by $[\varphi] = \{w \in W : w(\varphi) = 1\}$.

Then they consider the following four binary preference operators on propositions.

Definition 1 (*cf.* [1]) Given a preference model $\mathcal{M} = (\mathrm{W}, \preceq, \mathcal{V})$, one can define the following four binary preference operators on classical propositions:

- $\mathcal{M} \models \varphi \leq_{\exists\exists} \psi$ iff there exist $u \in [\varphi], v \in [\psi]$ such that $u \leq v$.

- $\mathcal{M} \models \varphi \leq_{\exists\forall} \psi$ iff there exists $u \in [\varphi]$, such that for all $v \in [\psi]$, $u \leq v$.
- $\mathcal{M} \models \varphi \leq_{\forall\exists} \psi$ iff for all $u \in [\varphi]$, there exists $v \in [\psi]$ such that $u \leq v$.
- $\mathcal{M} \models \varphi \leq_{\forall\forall} \psi$ iff for all $u \in [\varphi]$ and $v \in [\psi]$, then $u \leq v$.

Notice that these definitions of the truth conditions for the four preference operators can be interpreted as defining corresponding preference relations on $\mathcal{P}(\mathrm{W})$, the power set of W (which contains the sets $[\varphi]$) arising from a preorder on the set of worlds W. One can furthermore define two more preference operators on propositions:

- $\mathcal{M} \models \varphi \leq_{\exists\forall 2} \psi$ iff there exists $v \in [\psi]$, such that for all $u \in [\varphi]$, $u \leq v$
- $\mathcal{M} \models \varphi \leq_{\forall\exists 2} \psi$ iff for all $v \in [\psi]$, there exists $u \in [\varphi]$ such that $u \leq v$.

Therefore, from a given preorder on W we can consider six relations on subsets of W. The basic set-inclusions between these relations are given in the following proposition.

Proposition 1 *The following inclusions hold:*

$$\leq_{\forall\forall} \subseteq \leq_{\forall\exists} \subseteq \leq_{\exists\exists}, \quad \leq_{\forall\forall} \subseteq \leq_{\exists\forall} \subseteq \leq_{\exists\exists}, \quad \leq_{\forall\forall} \subseteq \leq_{\forall\exists 2} \subseteq \leq_{\exists\exists}, \quad \leq_{\forall\forall} \subseteq \leq_{\exists\forall 2} \subseteq \leq_{\exists\exists}$$

Moreover, the four intermediate relations are not comparable, except for the following inclusions:

$$\leq_{\exists\forall 2} \subseteq \leq_{\forall\exists}, \quad \leq_{\exists\forall} \subseteq \leq_{\forall\exists 2}.$$

Proof All the inclusion relations are easy to check. Moreover, the inclusions given in Proposition 1 are the only ones that are valid among the four intermediate relations, as the following examples show: Take $W = \{u_1, u_2, u_3, u_4, u_5, u_6\}$ and $A = \{u_1, u_2, u_3\}, B = \{u_4, u_5, u_6\}$. Then,

- If the preorder is defined by reflexivity plus $u_1 \leq u_4, u_2 \leq u_5$ and $u_3 \leq u_5$, then $A \leq_{\forall\exists} B$ is the unique intermediate relation that is satisfied.
- If the preorder is defined by reflexivity plus $u_1 \leq u_4, u_2 \leq u_5$ and $u_2 \leq u_6$, then $A \leq_{\forall\exists 2} B$ is the unique intermediate relation that is satisfied.
- If the preorder is defined by reflexivity plus $u_2 \leq u_4, u_2 \leq u_5$ and $u_2 \leq u_6$, then $A \leq_{\exists\forall} B$ and $A \leq_{\forall\exists 2} B$ are the unique intermediate relations that are satisfied.
- If the preorder is defined by reflexivity plus $u_1 \leq u_4, u_2 \leq u_4$ and $u_3 \leq u_4$, then $A \leq_{\exists\forall 2} B$ and $A \leq_{\forall\exists} B$ are the unique intermediate relations that are satisfied.

2.2 *The Fuzzy Preorder Case*

Now we study the case when $\leq$ is a fuzzy $\odot$-preorder on W, i.e., $\leq : \mathrm{W} \times \mathrm{W} \longrightarrow [0, 1]$ satisfying reflexivity ($[u \leq u] = 1$ for all $u \in \mathrm{W}$) and $\odot$-transitivity with respect to a given t-norm $\odot$ (for all $u, v, w \in \mathrm{W}$, $([u \leq v] \odot [v \leq w]) \leq [u \leq w]$), where $[u \leq v]$ denotes the value in $[0, 1]$ of the fuzzy relation $\leq$ applied to the ordered pair of elements $u, v \in \mathrm{W}$. We will assume that W is a *finite* set, and we will denote by δ_u the singleton $\{u\}$.

Generalising the classical case, we can define the following fuzzy relations on $\mathcal{P}(\mathrm{W})$ from a fuzzy preorder on W.

Definition 2 Given a fuzzy preorder $\leq$ on W, we can define the following six fuzzy relations on $\mathcal{P}(\mathrm{W})$. For any $A, B \in \mathcal{P}(\mathrm{W})$ we let:

- $[A \leq_{\exists\exists} B] = \sup_{u\in A} \sup_{v\in B} [u \leq v]$
- $[A \leq_{\exists\forall} B] = \sup_{u\in A} \inf_{v\in B} [u \leq v]$
- $[A \leq_{\forall\exists} B] = \inf_{u\in A} \sup_{v\in B} [u \leq v]$
- $[A \leq_{\forall\forall} B] = \inf_{u\in A} \inf_{v\in B} [u \leq v]$
- $[A \leq_{\forall\exists 2} B] = \inf_{v\in B} \sup_{u\in A} [u \leq v]$
- $[A \leq_{\exists\forall 2} B] = \sup_{v\in B} \inf_{u\in A} [u \leq v]$.

where the value of $A \leq_{\circ} B$ is denoted by $[A \leq_{\circ} B]$ with $\leq_{\circ}$ being anyone of the six relations.

It is clear that, since the preorder $\leq$ is valued on $[0, 1]$, these relations are also $[0, 1]$-valued. For each $a \in (0, 1]$, we will write $A \leq^{a}_{\exists\exists} B$ when $[A \leq_{\exists\exists} B] \geq a$ and analogously for the other relations.

Proposition 2 *For any sets $A, B \in \mathcal{P}(\mathrm{W})$, we have:*

- $[A \leq_{\forall\forall} B] \leq [A \leq_{\forall\exists} B] \leq [A \leq_{\exists\exists} B]$,
- $[A \leq_{\forall\forall} B] \leq [A \leq_{\forall\exists 2} B] \leq [A \leq_{\exists\exists} B]$,
- $[A \leq_{\forall\forall} B] \leq [A \leq_{\exists\forall} B] \leq [A \leq_{\exists\exists} B]$*, and*
- $[A \leq_{\forall\forall} B] \leq [A \leq_{\exists\forall 2} B] \leq [A \leq_{\exists\exists} B]$.

Moreover the four intermediate relations are not comparable, except for the same two cases (now inequalities) of Proposition 1.

Proof Analogous to the proof of Proposition 1.

Out of the above six possibilities, we will mainly focus on two of them, $\leq_{\forall\exists}$ and $\leq_{\forall\exists 2}$, in the rest of the paper. These are well-behaved extensions of an initial fuzzy $\odot$-preorder to model a weak preference relation on subsets, since in particular they keep being $\odot$-preorders. Moreover, combining them, we can capture a very natural (preference) ordering related to orderings of intervals. Indeed, suppose $(W, \leq)$ is a totally (classical) pre-ordered set, and we want to extend $\leq$ to an ordering on the set *Int*(W) of intervals of W. The two most usual ways to do this are the following:

(i) $[a, b] \leq_1 [c, d]$ if $a \leq c$ and $b \leq d$,
(ii) $[a, b] \leq_2 [c, d]$ if $b \leq c$.

The relation $\leq_1$ is considered for example in [2], and it turns out to be definable as the intersection of the $\leq_{\forall\exists}$ and $\leq_{\forall\exists 2}$ relations on *Int*(W), that is, $\leq_1 = \leq_{\forall\exists} \cap \leq_{\forall\exists 2}$, while the second, $\leq_2$, coincides with the (crisp) relation $\leq_{\forall\forall}$ on *Int*(A). Actually, $\leq_{\forall\forall}$ is not a preorder because it is only reflexive for singletons, but it is enough for our purposes. In next sections, we will study in the fuzzy case these three basic relations ($\leq_{\forall\exists}$, $\leq_{\forall\exists 2}$, $\leq_{\forall,\forall}$) on $\mathcal{P}(\mathrm{W})$ arising from a fuzzy preorder $\leq$ on W.

3 Characterizing the Relations $\leq_{\forall\exists}$, $\leq_{\forall\exists 2}$ and $\leq_{\forall\forall}$

The following propositions describe the main properties satisfied by each one of these relations. In what follows, we assume a given a fuzzy $\odot$-preorder $\leq$ on W and the fuzzy relations $\leq_{\forall\exists}$, $\leq_{\forall\exists 2}$ and $\leq_{\forall\forall}$ which are defined as in Definition 2.

Proposition 3 *The relation $\leq_{\forall\exists}$ satisfies the following properties, for all $A, B, C \in \mathcal{P}(\mathrm{W})$:*

1. *Inclusion:* $[A \leq_{\forall\exists} B] = 1$, *if* $A \subseteq B$
2. *$\odot$-Transitivity:* $[A \leq_{\forall\exists} B] \odot [B \leq_{\forall\exists} C] \leq [A \leq_{\forall\exists} C]$
3. *Left-OR:* $[(A \cup B) \leq_{\forall\exists} C] = \min([A \leq_{\forall\exists} C], [B \leq_{\forall\exists} C])$
4. *Restricted Right-OR:* $[A \leq_{\forall\exists} (B \cup C)] \geq \max([A \leq_{\forall\exists} B], [A \leq_{\forall\exists} C])$. *The inequality becomes an equality if A is a singleton.*

Proposition 4 *The relation $\leq_{\forall\exists 2}$ satisfies the following properties, for all $A, B, C \in \mathcal{P}(\mathrm{W})$:*

1. *Inclusion:* $[A \leq_{\forall\exists 2} B] = 1$, *if* $B \subseteq A$
2. *$\odot$-Transitivity:* $[A \leq_{\forall\exists 2} B] \odot [B \leq_{\forall\exists 2} C] \leq [A \leq_{\forall\exists 2} C]$
3. *Restricted Left-OR:* $[(A \cup B) \leq_{\forall\exists 2} C] \geq \max([A \leq_{\forall\exists 2} C], [B \leq_{\forall\exists 2} C])$. *The inequality becomes an equality if C is a singleton.*
4. *Right-OR:* $[A \leq_{\forall\exists 2} (B \cup C)] = \min([A \leq_{\forall\exists 2} B], [A \leq_{\forall\exists 2} C])$.

Proposition 5 *The relation $\leq_{\forall\forall}$ satisfies the following properties, for all $A, B, C \in \mathcal{P}(\mathrm{W})$:*

1. *Restricted reflexivity:* $[A \leq_{\forall\forall} A] = 1$ *iff A is a singleton*
2. *$\odot$-Transitivity:* $[A \leq_{\forall\forall} B] \odot [B \leq_{\forall\forall} C] \leq [A \leq_{\forall\forall} C]$
3. *Left-OR:* $[(A \cup B) \leq_{\forall\forall} C] = \min([A \leq_{\forall\forall} C], [B \leq_{\forall\forall} C])$
4. *Right-OR:* $[A \leq_{\forall\forall} (B \cup C)] = \min([A \leq_{\forall\forall} B], [A \leq_{\forall\forall} C])$
5. *Inclusions:* $[A \leq_{\forall\forall} B] \leq [A' \leq_{\forall\forall} B']$, *if* $A' \subseteq A, B' \subseteq B$.

The proofs of the these propositions are easy and we omit them. Observe that, as already mentioned above, $\leq_{\forall\forall}$ is not reflexive.

Actually, the properties given above fully characterize the different relations on $\mathcal{P}(\mathrm{W})$ as showed in the next theorem.

Theorem 1 *The following characterizations hold:*

(i) *Let $\leq_{AE}$ be a relation between sets of $\mathcal{P}(\mathrm{W})$ satisfying Properties 1, 2, 3 and 4 of Proposition 3. Then there exists a fuzzy $\odot$-preorder $\leq$ on the set W such that $\leq_{AE}$ coincides with $\leq_{\forall\exists}$ as defined in Definition 2.*

(ii) *Let $\leq_{AE2}$ be a relation between sets of $\mathcal{P}(\mathrm{W})$ satisfying Properties 1, 2, 3 and 4 of Proposition 4. Then there exists a fuzzy $\odot$-preorder $\leq$ on the set W such that $\leq_{AE2}$ coincides with $\leq_{\exists\forall 2}$ as defined in Definition 2.*

(iii) *Let $\leq_{AA}$ be a relation between sets of $\mathcal{P}(\mathrm{W})$ satisfying Properties 1, 2, 3, 4 and 5 of Proposition 5. Then there exists a fuzzy $\odot$-preorder $\leq$ on the set W such that $\leq_{AA}$ coincides with $\leq_{\forall\forall}$ as defined in Definition 2.*

Proof We show the case of $\preceq_{AE}$, the rest of cases are proved in a similar way. Define a relation on W by $[u \leq v] = [\delta_u \preceq_{AE} \delta_v]$. Clearly $\leq$ is a fuzzy preorder on W. Now take into account that, for all $A \in \mathcal{P}(\mathrm{W})$, $A = \bigcup\{\delta_u : u \in A\}$ and, applying Properties 3 and 4, it is obvious that for all $A, B \in \mathrm{W}$, then $[A \preceq_{AE} B] = \inf_{u \in A} \sup_{v \in B} [\delta_u \preceq \delta_v]$. Thus (i) is proved. ☐

4 Characterizing Strict Fuzzy Orders Associated to Fuzzy Preorders

It is well known that any crisp preorder $\leq$ on an universe W induces an equivalence (or indifference) relation $\equiv$ and an strict order $<$, defined as follows:

- $x \equiv y$ iff $x \leq y$ and $y \leq x$,
- $x < y$ iff $x \leq y$ and $x \neq y$ or, alternatively iff $x \leq y$ and $y \nleq x$.

Observe that these relations satisfy that $x \leq y$ iff either $x \equiv y$ or $x < y$. We will use this condition to define an strict fuzzy order associated to a fuzzy preorder.

In the fuzzy setting (see for example [2, 14]), from a fuzzy $\odot$-preorder $\leq : W \times W \to [0, 1]$ we can define:

- the maximal indistinguishability relation $v \equiv w$ contained in the fuzzy preorder, defined by $[x \equiv y] = [x \leq y] \wedge [y \leq x]$;
- the minimal strict fuzzy $\odot$-order $<$ that satisfies the following equation

$$[x \leq y] = [x < y] \oplus [x \equiv y] \tag{1}$$

where $\oplus$ is a T-conorm (for example the maximum or the bounded sum).

So defined, the relation $\equiv$ is reflexive, symmetric and $\odot$-transitive, and thus it is a $\odot$-indistinguishability relation (the generalization of the crisp notion of equivalence relation). On the other hand, the minimal solution for b of the equation $a \leq b \oplus c$ in $[0, 1]$, is the so-called dual resituated implication, or implication associated to the T-conorm $\oplus$, which is defined as, It should be:

$$c \to^{\oplus} a = \inf\{b \mid c \oplus b \geq a\}.$$

Therefore, we take as the strict fuzzy order relation $<$ associated to $\leq$ for the T-conorm $\oplus$ the fuzzy relation defined as

$$[x < y] = [x \equiv y] \to^{\oplus} [x \leq y] = [y \leq x] \to^{\oplus} [x \leq y].$$

An easy computation shows that the strict fuzzy order relation for $\oplus = \max$ is defined as

$$[x < y] = \begin{cases} [x \leq y], \text{ if } [x \leq y] > [y \leq x], \\ 0, \quad \text{otherwise.} \end{cases} \tag{2}$$

And for $\oplus$ being the bounded sum (i.e. the Łukasiewicz T-conorm) is[1]

$$[x < y] = \begin{cases} [x \leq y] - [y \leq x], & \text{if } [x \leq y] > [y \leq x], \\ 0, & \text{otherwise.} \end{cases}$$

The strict relation associated to $\leq$ is a irreflexive ($[x < x] = 0$) and antisymmetric ($\min([x < y], [y < x]) = 0$) but, as far as we know, it is not known whether it is $\odot$-transitive in general. Nevertheless we have the following result.

Proposition 6 *Let $\leq$ be a* min*-preorder on a universe* W *and let $<$ be the associated strict relation w.r.t.* $\oplus = \max$. *Then $<$ is* min*-transitive.*

Proof The proof is by contradiction. Suppose the strict relation is not min-transitive. Then there must exist elements $u, v, w \in W$ such that $[u < v], [v, w] > 0$ and $[u < w] = 0$ which is equivalent that $[u \leq v] = a > b = [v \leq u]$, $[v \leq w] = c > d = [w \leq v]$ and $[u \leq w] = [w \leq u] = f$. Thus we have five values a, b, c, d, f and we know that

$$a > b \text{ and } c > d. \tag{*}$$

We can now reason by cases:

(1) Suppose $a \geq c$ and $b \geq d$. Combining this assumption with (*) we have that $a \geq c > d$. By transitivity, $f \geq \min(a, c) = c$ and $f \geq \min(d, b) = d$ by hypothesis. Moreover $\min([w \leq u], [u \leq v]) = \min(f, a) \leq d = [w \leq v]$. This implies that $a \leq d$, in contradiction with the fact that $d < a$.
(2) Suppose $a \geq c$ and $b < d$. Combining this assumption with (*) we have that $d < c \leq a$. By transitivity, $f \geq \min(a, c) = c$ and $f \geq \min(d, b) = b$ by hypothesis. Moreover $\min([w \leq u], [u \leq v]) = \min(f, a) \leq d = [w \leq v]$. This implies that $f \leq d$, and by hypothesis $f \leq d < c$, in contradiction with $f \geq c$ previously proved.
(3) Suppose $a \leq c$ and $b \geq d$. Combining this assumption with (*) we have that $b < a \leq c$. By transitivity, $f \geq \min(a, c) = a$ and $f \geq \min(d, b) = d$ by hypothesis. Moreover $\min([v \leq w], [w \leq u]) = \min(c, f) \leq b = [v \leq u]$. This imply that $f \leq b$ and by hypothesis $f \leq b < a$, in contradiction with $f \geq a$ previously proved.
(4) Suppose $a \leq c$ and $b \leq d$. Combining this assumption with (*) we have that $b < a \leq c$. By transitivity, $f \geq \min(a, c) = a$ and $f \geq \min(d, b) = b$ by hypothesis. Moreover $\min([v \leq w], [w \leq u]) = \min(c, f) \leq b = [v \leq u]$. This implies that $f \leq b$, and by hypothesis $f \leq b < a$, in contradiction with $f \geq a$ previously proved. □

From now on, we consider the strict fuzzy order $<$ associated to $\leq$ the one defined by taking $\oplus = \max$ according to (2).

Now we can come back to the topic of how to define a strict fuzzy order relation on sets of $\mathcal{P}$(W) corresponding to a fuzzy preorder in W. Halpern notices in [15] that

[1] This is the strict order companion defined and studied in [7].

there are two different ways to define (in the crisp case) a strict relation on $\mathcal{P}(\mathrm{W})$ from a preorder on W. The same idea applied to the fuzzy case gives rise to the following two possible definitions for the strict relations:

- The *standard method*, that amounts to define

$$[A <_\circ B] = \begin{cases} [A \leq_\circ B], & \text{if } [A \leq_\circ B] > [B \leq_\circ A] \\ 0, & \text{otherwise.} \end{cases}$$

  This means in fact to use (2) to define $[A <_\circ B]$ as the value of the strict order associated to the preorder $\leq_\circ$, where $\leq_\circ$ is either $\leq_{\forall\exists}$, $\leq_{\forall\exists 2}$ or $\leq_{\forall\forall}$.
- The *alternative method*, that first considers the strict order $<$ on companion of $\leq$ in W according to (2), and then defines $<_{\forall\exists}$, $<_{\forall\exists 2}$ and $<_{\forall\forall}$ on $\mathcal{P}(\mathrm{W})$ according to Definition 2, but replacing $\leq$ by $<$.

In general, these two methods give rise to *two different* irreflexive and (restricted) antisymmetric strict relations as the following examples show:

Example 1 Consider the $\forall\exists$ extension. Notice first that the alternative method gives

$$[A <_{\forall\exists} B] = \inf_{u \in A} \sup_{v \in B} [u < v].$$

The counterexample is the following. Take the four element set $W = \{u_1, u_2, u_3, u_4\}$, with the following fuzzy preorder: reflexivity ($[x \leq x] = 1$) plus $[u_1 \leq u_3] = [u_3 \leq u_1] = a$ and $[u_2 \leq u_4] = b$, with $a, b \neq 0$. The associated strict relation on W is the one having only one pair of elements with value different from 0. Indeed an easy computation shows that $[u_2 < u_4] = b$. Let $A = \{u_1, u_2\}$ and let $B = \{u_3, u_4\}$. Then:

- To compute the value of $[A <_{\forall\exists} B]$ according to the standard method, we have to compute first: $[A \leq_{\forall\exists} B] = ([u_1 \leq u_3] \vee [u_1 \leq u_4]) \wedge ([u_2 \leq u_3] \vee [u_2 \leq u_4]) = a \wedge b \neq 0$, $\quad [B \leq_{\forall\exists} A] = ([u_3 \leq u_1] \vee [u_3 \leq u_2]) \wedge ([u_4 \leq u_1] \vee [u_4 \leq u_2]) = a \wedge 0 = 0$. Then, by definition, we have $[A <_{\forall\exists} B] = a \wedge b \neq 0$.
- With the alternative method, the value of $[A <_{\forall\exists} B]$ is computed as $[A <_{\forall\exists} B] = \inf_{u \in A} \sup_{v \in B} [u < v] = 0$.

Therefore the obtained strict relations are different. □

Example 2 Consider now the $\forall\forall$ extension. Take $\mathrm{W} = \{w_1, w_2\}$ with the preorder $[w_1 \leq w_1] = [w_2 \leq w_2] = [w_1 \leq w_2] = 1$ and $[w_2 \leq w_1] = 0$. Further, take $A = \{w_1\}$ and $B = \mathrm{W}$. Then it is obvious that $[A \leq_{\forall\forall} B] = 1$ and $[B \leq_{\forall\forall} A] = 0$. Therefore, according to the standard method, we have $[A <_{\forall\forall} B] = 1$, while according to the alternative method, we have $[A <_{\forall\forall} B] = \inf_{u \in A} \inf_{v \in B} [u < v] = 0$. □

Notice that strict relations obtained by the alternative method are $\odot$-transitive (so they are strict orders), but this is not clear for strict relations obtained by the standard method. In fact we have the following open problems:

- Let $\leq$ be a strictly monotonic fuzzy preorder on W and let $\leq_\circ$ be one of the fuzzy preorders defined on $\mathcal{P}(\mathrm{W})$ considered in the previous sections. Is the strict relation obtained by the standard method $\odot$-transitive?
- Is there some order relation between the strict orders obtained by the standard and the alternative methods?
- It is obvious that the strict order $<$ on W and the strict order on $\mathcal{P}(\mathrm{W})$ obtained from the preorder by the standard method satisfy the following anti-symmetry property: for all $A, B \in \mathcal{P}(\mathrm{W})$, $\min([A <_\circ B], [B <_\circ A]) = 0)$. It is clear that for singletons the strict order obtained by the alternative method satisfies the same anti-symmetry property but, is this true for the strict order obtained by the alternative method in general? Otherwise, what type of anti-symmetry property does it satisfy?

Therefore, taking into account that we are interested in obtaining strict fuzzy orders (irreflexive and $\odot$-transitive relations), in the rest of the paper we will consider the strict relations obtained by the *alternative method* and its characteristics properties. Next theorem provides a characterization result for these strict orders.

Theorem 2 *The following characterizations hold:*

(i) *Let $\prec_{AE}$ be a relation between sets of $\mathcal{P}(\mathrm{W})$ satisfying Properties 2, 3 and 4 of Proposition 3 plus irreflexivity ($[A \prec_{AE} A] = 0$) and restricted anti-symmetry ($\min([A \prec_{AE} B], [B \prec_{AE} A]) = 0$ for all singletons $A, B \in \mathcal{P}(\mathrm{W})$). Then there exists a fuzzy $\odot$-preorder $\leq$ on the set W such that $\prec_{AE} = <_{\forall\exists}$.*

(ii) *Let $\prec_{AE2}$ be a relation between sets of $\mathcal{P}(\mathrm{W})$ satisfying Properties 2, 3 and 4 of Proposition 4 plus irreflexivity and restricted anti-symmetry. Then there exists a fuzzy $\odot$-preorder $\leq$ on the set W such that $\prec_{AE2} = <_{\exists\forall 2}$.*

(iii) *Let $\prec_{AA}$ be a relations between sets of $\mathcal{P}(\mathrm{W})$ satisfying Properties 2, 3, 4 and 5 of Proposition 5 plus irreflexivity and anti-symmetry. Then there exists a fuzzy $\odot$-preorder $\leq$ on the set W such that $\prec_{AA} = <_{\forall\forall}$.*

At the end of Sect. 2.2 we mentioned that one of the preorders we were interested in was the (crisp) relation $\leq_1$, whose fuzzy counterpart can be defined by $[x \leq_1 y] = \min([x \leq_{\forall\exists} y], [x \leq_{\forall\exists 2} y])$. Consequently, in Sect. 3 we separately characterized the fuzzy preorders $\leq_{\forall\exists}$ and $\leq_{\forall\exists 2}$, and the same is applicable to the corresponding strict orders studied in this section. We will move now to a logical approach to preference relations and to the previously studied fuzzy relations. In particular, in the next section we study a logical setting to reason about fuzzy preferences on classical propositions by means of several two-tiered modal logics, with binary modal operators corresponding to fuzzy preorders and strict orders separately, and after we show the desired preorder and strict order are definable in a yet another modal logic combining the previous ones.

5 A Modal Logic to Reason About Preferences

In this section three logics to reason about conditional (syntactic) objects capturing the idea of the preference relations $\preceq_\circ$ (for $\circ \in \{\forall\exists, \forall\exists 2, \forall\forall\}$) are defined and studied, using similar techniques from [12].

Throughout this section, in order to simplify matters, rather than defining the logic relative to preference degrees in [0, 1] and a t-norm, we will restrict ourselves to deal with a totally ordered finite set V of preference degrees (with 1 and 0 as its top and bottom elements), and we will fix a *finite t-norm* $\odot$ (see e.g. [19]) on V.

On these grounds, we define, model-theoretically, a common framework for several logics of preference relations, LAP for short, as follows.

Definition 3 The language of LAP is two levelled:

- The first level ($\mathcal{L}_0$ language) contains *propositional formulas* of LAP that are built up from a finite set of *variables* $Var = \{p_1, \ldots, p_N\}$ and the *constants* $\perp, \top$ by means of the binary operators $\wedge$ and $\vee$ and the unary operator $\neg$. The set of propositional formulas is denoted by $\mathcal{P}$.
- The second level ($\mathcal{L}_1$ language) contains:
  - *Atomic graded preference* formulas of LAP that are triples

$$\varphi \preceq^a \psi$$

    consisting of two propositional formulas φ and ψ from $\mathcal{L}_0$, and a value $a \in V \setminus \{0\}$.
  - (General) *preference formulas* of LAP are built up from atomic graded preferences and the constants $\perp, \top$ by means of the binary connectives $\wedge$ and $\vee$ and the unary connective $\neg$.

The semantics is given by $\odot$-preference Kripke models $\mathcal{M} = (W, \preceq, e)$ where W is a finite set of worlds, $\preceq$: $\mathrm{W} \times \mathrm{W} \to V$ is a $\odot$-fuzzy preorder relation, and e : $\mathrm{W} \times Var \mapsto \{0, 1\}$ is a Boolean evaluation of propositional variables in every world, which is extended to propositions of $\mathcal{L}_0$ in the usual way for classical propositions. For each $\mathcal{L}_0$-proposition φ, we will denote by $[\varphi]_\mathcal{M}$ the set $\{w \in \mathrm{W} : e(w, \varphi) = 1\}$ of worlds satisfying φ.

For each $\circ \in \{\forall\exists, \forall\exists 2, \forall\forall\}$, each Kripke model $\mathcal{M} = (\mathrm{W}, S, e)$ induces a Boolean truth-evaluation of $\mathcal{L}_1$-formulas $e^\circ_\mathcal{M} : \mathcal{L}_1 \to \{0, 1\}$ defined as follows:

- for atomic preference formulas: $e^\circ_\mathcal{M}(\varphi \preceq^a \psi) = 1$ if $[[\varphi]_\mathcal{M} \preceq_\circ [\psi]_\mathcal{M}] \geq a$, and $e_\mathcal{M}(\varphi \preceq^a \psi) = 0$ otherwise.
- for compound formulas, use the usual Boolean truth functions.

From there, we can define the notion of logical consequence in the logic LAP for preference formulas.

Definition 4 Let $\circ \in \{\forall\exists, \forall\exists 2, \forall\forall\}$. Let $T \cup \{\Phi\}$ be a set of preference formulas. We say that Φ logically follows from T under the $\preceq_\circ$ semantics, written $T \models^\circ_{\mathsf{LAP}}$

Φ, if for every Kripke model $\mathcal{M} = (\mathrm{W}, \leq, e)$, if $e^{\circ}_{\mathcal{M}}(\Psi) = 1$ for every $\Psi \in T$, then $e^{\circ}_{\mathcal{M}}(\Phi) = 1$ as well.

In the following, for every Boolean evaluation ω of the propositional variables *Var*, we will denote by $\overline{\omega}$ the maximally elementary conjunction (m.e.c. for short) of all the N literals made true by ω. Obviously, every proposition is logically equivalent to a disjunction of m.e.c.'s.

Next subsections are devoted to the axiomatization of the particular logics for $\leq_{\forall\exists}$, $\leq_{\forall\exists 2}$ and $\leq_{\forall\forall}$.

5.1 The Logic $\mathsf{LAP}_{\forall\exists}$ Corresponding to the $\leq_{\forall\exists}$ Preference Relation

Recall that, when $\circ = \forall\exists$, the semantics we have in each Kripke model $\mathcal{M}$ is:

$$e_{\mathcal{M}}(\varphi \preceq^{a} \psi) = 1 \text{ iff } [[\varphi]_{\mathcal{M}} \leq_{\forall\exists} [\psi]_{\mathcal{M}}] = (\inf_{u \in [\varphi]_{\mathcal{M}}} \sup_{w \in [\psi]_{\mathcal{M}}} [u \leq w]) \geq a.$$

Building on this intended semantics, we propose the following axiomatization of $\mathsf{LAP}_{\forall\exists}$.

Definition 5 The following are the axioms for $\mathsf{LAP}_{\forall\exists}$:

(A1)	Axioms of classical propositional calculus (CPC) for $\mathcal{L}_1$-formulas	
(A2)	$\varphi \preceq^{1} \psi$, where $\varphi \to \psi$ is a tautology of CPC	
(A3)	$(\varphi \preceq^{a} \psi) \wedge (\psi \preceq^{b} \chi) \to (\varphi \preceq^{a \odot b} \chi)$	(transitivity)
(A4)	$(\varphi \preceq^{a} \psi) \to (\varphi \preceq^{b} \psi)$, where $a \leq b$	(nestedness)
(A5)	$(\varphi \vee \psi \preceq^{a} \chi) \leftrightarrow (\varphi \preceq^{a} \chi) \wedge (\psi \preceq^{a} \chi)$	(Left-OR)
(A6)	$(\overline{\omega} \preceq^{a} \varphi \vee \psi) \leftrightarrow (\overline{\omega} \preceq^{a} \varphi) \vee (\overline{\omega} \preceq^{a} \psi)$	(restricted Right-OR)

The only rule of $\mathsf{LAP}_{\forall\exists}$ is Modus Ponens.

We will denote by $\vdash^{\forall\exists}_{\mathsf{LAP}}$ the notion of deduction relative to the axiomatic system just defined.

Theorem 3 *For any set $T \cup \{\Phi\}$ of $\mathcal{L}_1$-formulas, it holds that $T \models^{\forall\exists}_{\mathsf{LAP}} \Phi$ if, and only if, $T \vdash^{\forall\exists}_{\mathsf{LAP}} \Phi$.*

Proof One direction is soundness, and it is an easy computation, see Proposition 3. As for the other direction, assume $T \nvdash^{\forall\exists}_{\mathsf{LAP}} \Phi$. The idea is to consider the graded expressions $\varphi \preceq^{a} \psi$ as propositional (Boolean) variables that are ruled by the axioms together with the laws of classical propositional logic CPC. Let Γ be the set of all possible instantiations of axioms (A1)–(A6). Then it holds that Φ does not follow from $T \cup \Gamma$ using CPC reasoning, i.e. $T \cup \Gamma \nvdash_{CPC} \Phi$. By completeness of CPC, there exists a Boolean interpretation v such that $v(\Psi) = 1$ for all $\Psi \in T \cup \Gamma$ and

$v(\Phi) = 0$. Now we will build a $\odot$-preference Kripke model $\mathcal{M}$ such that $e_{\mathcal{M}}(\Psi) = 1$ for all $\Psi \in T$ and $e_{\mathcal{M}}(\Phi) = 0$. To do that, we take $\Omega = \{\omega : Var \longrightarrow \{0, 1\}\}$, i.e. the set of interpretations of propositional language, and define $\leq : \Omega \times \Omega \to V$ by

$$[\omega \leq \omega'] = \max\{a \in V \mid v(\overline{\omega} \preceq^a \overline{\omega'}) = 1\}.$$

By axioms (A2), (A3), $\leq$ is a $\odot$-preorder. Consider the model $\mathcal{M} = (\Omega, \leq, e)$, where for each $\omega \in \Omega$ and $p \in Var$, $e(\omega, p) = \omega(p)$. What remains to check is that $e_{\mathcal{M}}(\Psi) = v(\Psi)$ for every $\mathsf{LAP}_{\forall\exists}$-formula Ψ. In order to prove this equality it suffices to show that, for every $\varphi, \psi \in \mathcal{L}_0$ and $a \in [0, 1]$, we have $e_{\mathcal{M}}(\varphi \preceq^a \psi) = v(\varphi \preceq^a \psi)$, that is, to prove that

$$v(\varphi \preceq^a \psi) = 1 \quad \text{iff} \quad \inf_{\omega \in [\varphi]_{\mathcal{M}}} \sup_{\omega' \in [\psi]_{\mathcal{M}}} [\omega \leq \omega'] \geq a.$$

By axioms (A5) and (A6), we have that $\mathsf{LAP}_{\forall\exists}$ proves

$$\varphi \preceq^a \psi \;\leftrightarrow\; \bigwedge_{\omega \in \Omega : \omega(\varphi) = 1} \; \bigvee_{\omega' \in \Omega : \omega'(\psi) = 1} \overline{\omega} \preceq^a \overline{\omega'}.$$

Therefore, $v(\varphi \preceq^a \psi) = 1$ iff for all $\omega \in \Omega$ such that $\omega(\varphi) = 1$, there exists $w' \in \Omega$ such that $\omega'(\psi) = 1$ and $v(\overline{\omega} \preceq^a \overline{\omega'}) = 1$. But, as we have previously observed, $v(\overline{\omega} \preceq^a \overline{\omega'}) = 1$ holds iff $[\omega \leq \omega'] \geq a$. In other words, we actually have $v(\varphi \preceq^a \psi) = 1$ iff $\min_{\omega \in [\varphi]_{\mathcal{M}}} \max_{\omega' \in [\psi]_{\mathcal{M}}} [\omega \leq \omega'] \geq a$. This concludes the proof. $\square$

5.2 The Logics $\mathsf{LAP}_{\forall\exists 2}$ and $\mathsf{LAP}_{\forall\forall}$ Corresponding to the $\leq_{\forall\exists 2}$ and $\leq_{\forall\forall}$ Preference Relations

In a very similar way, with the obvious changes, we can define axiomatic systems for the logics of $\mathsf{LAP}_{\forall\exists 2}$ and $\mathsf{LAP}_{\forall\forall}$. We do not include the completeness proofs since they are analogous to the one for $\mathsf{LAP}_{\forall\exists}$.

Recall that, under the $\forall\exists 2$ semantics, the evaluation of atomic preference formulas in a preference Kripke model $\mathcal{M}$ is as follows:

$$e_{\mathcal{M}}(\varphi \preceq^a \psi) = 1 \text{ iff } [[\varphi]_{\mathcal{M}} \leq_{\forall\exists 2} [\psi]_{\mathcal{M}}] = (\inf_{w \in [\psi]_{\mathcal{M}}} \sup_{u \in [\varphi]_{\mathcal{M}}} [u \leq w]) \geq a.$$

Theorem 4 *Let* $\mathsf{LAP}_{\forall\exists 2}$ *be the axiomatic system whose axioms are:*

(A1)	*Axioms of CPC for* $\mathcal{L}_1$*-formulas*	
(A2)	$\varphi \preceq^1 \psi$, *where* $\psi \to \varphi$ *is a tautology of CPC*	
(A3)	$(\varphi \preceq^a \psi) \wedge (\psi \preceq^b \chi) \to (\varphi \preceq^{a \odot b} \chi)$	*(transitivity)*
(A4)	$(\varphi \preceq^a \psi) \to (\varphi \preceq^b \psi)$, *for all* $a \leq b$	*(nestedness)*
(A5)	$(\varphi \preceq^a \psi \vee \chi) \;\leftrightarrow\; (\varphi \preceq^a \psi) \wedge (\varphi \preceq^a \chi)$	*(Right-OR)*

(A6) $(\varphi \vee \psi \preceq^a \overline{\omega}) \leftrightarrow (\varphi \preceq^a \overline{\omega}) \vee (\preceq^a \psi \preceq^a \overline{\omega})$ *(restricted Left-OR)*

and whose only inference rule is modus ponens. Then $\mathsf{LAP}_{\forall\exists 2}$ *is sound and complete with respect to the class of* $\odot$*-preference Kripke models under the* $\forall\exists 2$ *semantics.*

As for the $\forall\forall$ semantics, the evaluation of atomic preference formulas in a preference Kripke model $\mathcal{M}$ is:

$$e_{\mathcal{M}}(\varphi \preceq^a \psi) = 1 \text{ iff } [[\varphi]_{\mathcal{M}} \preceq_{\forall\forall} [\psi]_{\mathcal{M}}] = (\inf_{u \in [\varphi]_{\mathcal{M}}} \inf_{w \in [\psi]_{\mathcal{M}}} [u \leq w]) \geq a.$$

Theorem 5 *Let* $\mathsf{LAP}_{\forall\forall}$ *be the axiomatic system whose axioms are:*

(A1) *Axioms of CPC for* $\mathcal{L}_1$*-formulas*
(A2) $(\varphi \preceq^a \psi) \to (\varphi' \preceq^a \psi')$*, where* $\varphi' \to \varphi, \psi' \to \psi$ *are tautologies of CPC*
(A3) $\overline{\omega} \preceq^1 \overline{\omega}$ *(restricted reflexivity)*
(A4) $(\varphi \preceq^a \psi) \wedge (\psi \preceq^b \chi) \to (\varphi \preceq^{a \odot b} \chi)$ *(transitivity)*
(A5) $(\varphi \preceq^a \psi) \to (\varphi \preceq^b \psi)$*, for all* $a \leq b$ *(nestedness)*
(A6) $(\varphi \vee \psi \preceq^a \chi) \leftrightarrow (\varphi \preceq^a \chi) \wedge (\psi \preceq^a \chi)$ *(Left-OR)*
(A7) $(\psi \preceq^a \varphi \vee \psi) \leftrightarrow (\psi \preceq^a \varphi) \wedge (\psi \preceq^a \psi)$ *(Right-OR)*

and whose only inference rule is modus ponens. Then $\mathsf{LAP}_{\forall\forall}$ *is sound and complete with respect to the class of* $\odot$*-preference Kripke models under the* $\forall\forall$ *semantics.*

Moreover, in the same way, we could axiomatize the logics $\mathsf{LAP}^s_{\forall\exists}$, $\mathsf{LAP}^s_{\forall\exists 2}$ and $\mathsf{LAP}^s_{\forall\forall}$ corresponding to the associated strict preference orders.

Nevertheless our goal is to axiomatize the logic modeling preference triples $\langle \leq, <, \equiv \rangle$ corresponding to the preference relations $\preceq_1 = \preceq_{\forall\exists} \wedge \preceq_{\forall\exists 2}$ and $\preceq_2 = \preceq_{\forall\forall}$. The axiomatizations of these logics are given in the next section.

5.3 The Logic LAP^1

In this subsection we define and study the logic corresponding to the fuzzy preorder $\preceq_1 = \preceq_{\forall\exists} \wedge \preceq_{\forall\exists 2}$.

The language of LAP^1 is as the one for LAP with the difference that now we have four kinds of atomic preference formulas:

$$\varphi \preceq^a_\alpha \psi, \quad \varphi \preceq^a_\beta \psi, \quad \varphi \prec^a_\alpha \psi, \quad \varphi \prec^a_\beta \psi,$$

where $a \in V \setminus \{0\}$. The semantics is still given by $\odot$-preference Kripke models $\mathcal{M} = (\mathrm{W}, \leq, e)$, where $e_{\mathcal{M}}$ evaluates the above kinds of atomic preference formulas in the expected way:

- $e_{\mathcal{M}}(\varphi \preceq^a_\alpha \psi) = 1$ if $[[\varphi]_{\mathcal{M}} \preceq_{\forall\exists} [\psi]_{\mathcal{M}}] = (\inf_{u \in [\varphi]_{\mathcal{M}}} \sup_{w \in [\psi]_{\mathcal{M}}} [u \leq w]) \geq a$
- $e_{\mathcal{M}}(\varphi \preceq^a_\beta \psi) = 1$ if $[[\varphi]_{\mathcal{M}} \preceq_{\forall\exists 2} [\psi]_{\mathcal{M}}] = (\inf_{w \in [\psi]_{\mathcal{M}}} \sup_{u \in [\varphi]_{\mathcal{M}}} [u \leq w]) \geq a$
- $e_{\mathcal{M}}(\varphi \prec^a_\alpha \psi) = 1$ if $[[\varphi]_{\mathcal{M}} \prec_{\forall\exists} [\psi]_{\mathcal{M}}] = (\inf_{u \in [\varphi]_{\mathcal{M}}} \sup_{w \in [\psi]_{\mathcal{M}}} [u < w]) \geq a$

- $e_{\mathcal{M}}(\varphi \prec_{\beta}^{a} \psi) = 1$ if $[[\varphi]_{\mathcal{M}} <_{\forall\exists 2} [\psi]_{\mathcal{M}}] = (\inf_{w \in [\psi]_{\mathcal{M}}} \sup_{u \in [\varphi]_{\mathcal{M}}} [u < w]) \geq a$.

The notion of logical consequence is defined as usual, and will be denoted by $\models_{\mathsf{LAP}^1}$.

Next we propose an axiomatic system for LAP^1.

Definition 6 The axioms for LAP^1 are:

- Axioms of $\mathsf{LAP}_{\forall\exists}$ for the $\preceq_{\alpha}^{a}$ operators.
- Axioms of $\mathsf{LAP}_{\forall\exists 2}$ for the $\preceq_{\beta}^{a}$ operators.
- Axioms for the $\prec_{\alpha}^{a}$ operators:

(AS1)	$\neg(\varphi \prec_{\alpha}^{a} \varphi)$	(irreflexivity)
(AS2)	$\neg((\overline{\varpi} \prec_{\alpha}^{a} \overline{\omega'}) \wedge (\overline{\omega'} \prec_{\alpha}^{b} \overline{\varpi}))$	(restricted anti-symmetry)
(AS3)	$(\varphi \prec_{\alpha}^{a} \psi) \wedge (\psi \prec_{\alpha}^{b} \chi) \rightarrow (\varphi \prec_{\alpha}^{a*b} \chi)$	($\odot$-transitivity)
(AS4)	$(\varphi \prec_{\alpha}^{a} \psi) \rightarrow (\varphi \prec_{\alpha}^{b} \psi)$, for all $a \leq b$	(nestedness)
(AS5)	$(\varphi \prec_{\alpha}^{a} \overline{\varpi}) \wedge (\psi \prec_{\alpha}^{a} \overline{\varpi}) \leftrightarrow (\varphi \vee \psi \prec_{\alpha}^{a} \overline{\varpi})$	(Left-OR)
(AS6)	$(\chi \prec_{\alpha}^{a} \varphi \vee \psi) \leftrightarrow (\chi \prec_{\alpha}^{a} \varphi) \vee (\chi \prec_{\alpha}^{a} \psi)$	(Restricted Right-OR)

- Axioms for the $\prec_{\beta}^{a}$ operators:

(BS1)	$\neg(\varphi \prec_{\beta}^{a} \varphi)$	(irreflexivity)
(BS2)	$\neg((\overline{\varpi} \prec_{\beta}^{a} \overline{\omega'}) \wedge (\overline{\omega'} \prec_{\beta}^{b} \overline{\varpi}))$	(restricted anti-symmetry)
(BS3)	$(\varphi \prec_{\beta}^{a} \psi) \wedge (\psi \prec_{\beta}^{b} \chi) \rightarrow (\varphi \prec_{\beta}^{a*b} \chi)$	($\odot$-transitivity)
(BS4)	$(\varphi \prec_{\beta}^{a} \psi) \rightarrow (\varphi \prec_{\beta}^{b} \psi)$, for all $a \leq b$	(nestedness)
(BS5)	$(\varphi \vee \psi \prec_{\beta}^{a} \chi) \leftrightarrow (\varphi \prec_{\beta}^{a} \chi) \wedge (\psi \prec_{\beta}^{a} \chi)$	(Restricted Left-OR)
(BS6)	$(\overline{\varpi} \prec_{\beta}^{a} \varphi \vee \psi) \leftrightarrow (\overline{\varpi} \prec_{\beta}^{a} \varphi) \vee (\overline{\varpi} \prec_{\beta}^{a} \psi)$	(Right-OR)

- Connecting axioms:

(AB) $\overline{\varpi} \preceq_{\alpha}^{a} \overline{\omega'} \leftrightarrow \overline{\varpi} \preceq_{\beta}^{a} \overline{\omega'}$

(ABS) $\overline{\varpi} \prec_{\alpha}^{a} \overline{\omega'} \leftrightarrow \overline{\varpi} \prec_{\beta}^{a} \overline{\omega'}$

(SA1) $\bigwedge \left((\overline{\varpi} \preceq_{\alpha}^{a} \overline{\omega'}) \rightarrow (\overline{\omega'} \preceq_{\alpha}^{a} \overline{\varpi}) : a > 0 \right) \rightarrow \neg(\overline{\varpi} \prec_{\alpha}^{a_0} \overline{\omega'})$, where a_0 is the minimum element of $V \setminus \{0\}$.

(SA2) $\neg \bigwedge \left((\overline{\varpi} \preceq_{\alpha}^{a} \overline{\omega'}) \rightarrow (\overline{\omega'} \preceq_{\alpha}^{a} \overline{\varpi}) : a > 0 \right) \rightarrow \left((\overline{\varpi} \prec_{\alpha}^{b} \overline{\omega'}) \leftrightarrow (\overline{\varpi} \preceq_{\alpha}^{b} \overline{\omega'}) \right)$

The only inference rule for LAP^1 is Modus Ponens.

Observe that axiom (AB) is related to the fact that (semantically), over m.e.c.'s, the weak relations $\preceq_{\alpha}$ and $\preceq_{\beta}$ coincide, and the same for axiom (ABS) regarding the strict relations $\prec_{\alpha}$ and $\prec_{\beta}$. Finally, axioms (SA1) and (SA2) are for $\prec_{\alpha}$ a logical translation of the definition of strict order $<$ from the preorder $\leq$ on W according to Eq. (2). Note that analogous axioms for $\prec_{\beta}$ can be derived using axiom (AB).

Denoting by $\vdash_{\mathsf{LAP}^1}$ the notion of proof in LAP^1, we have the following completeness result.

Theorem 6 *For any set $T \cup \{\Phi\}$ of $\mathcal{L}_1$-formulas, it holds that $T \models_{\mathsf{LAP}^1} \Phi$ if, and only if, $T \vdash_{\mathsf{LAP}^1} \Phi$.*

Proof One direction is soundness. Let $M = (W, \leq, e)$ a $\odot$-preference Kripke model. Axiom (AB) holds in M since both preorders $\leq_{\forall\exists}$ and $\leq_{\forall\exists 2}$ are defined from the same preorder $\leq$ on W, and thus they coincide over the singletons. The same argument is valid for (ABS), exchanging preorder by strict order. Axioms (SA1) and (SA2) correspond to the definition of the strict order $<$ on W from the preorder $\leq$. The interpretation of (AS1) roughly says that, for elements of W, if $[u \leq v] \leq [v \leq u]$ then $[u < v] = 0$ and (AS2) says that if $[u \leq v] > [v \leq u]$ then $[u < v] = [u \leq v]$.

As for the converse direction, assume $T \not\vdash_{\mathsf{LAP}^1} \Phi$. The construction of the countermodel is very similar to that of Theorem 3, and the idea is again to consider all atomic preference formulas $\varphi \lhd^a \psi$ (with $\lhd \in \{\preceq_\alpha, \preceq_\beta, \prec_\alpha, \prec_\beta\}$) as propositional (Boolean) variables that are ruled by the laws of classical propositional logic CPC. Let Γ be the set of all possible instantiations of axioms of LAP^1. Then it follows that Φ does not follow from $T \cup \Gamma$ using CPC reasoning, i.e. $T \cup \Gamma \not\vdash_{CPC} \Phi$. By completeness of CPC, there exists a Boolean interpretation v such that $v(\Psi) = 1$ for all $\Psi \in T \cup \Gamma$ and $v(\Phi) = 0$. Now we will build a $\odot$-preference Kripke model $\mathcal{M} = (\Omega, \leq, e)$ such that $e_{\mathcal{M}}(\Psi) = 1$ for all $\Psi \in T$ and $e_{\mathcal{M}}(\Phi) = 0$. We take $\Omega = \{\omega : Var \longrightarrow \{0, 1\}\}$, i.e. the set of Boolean interpretations of propositional variables, and define $\leq \colon \Omega \times \Omega \to [0, 1]$ by

$$[\omega \leq \omega'] = \max\{a \in V \mid v(\overline{\omega} \preceq^a_\alpha \overline{\omega'}) = 1\}.$$

Notice that, by axiom(AB), this value is equal to $\max\{a \in V \mid v(\overline{\omega} \preceq^a_\beta \overline{\omega'}) = 1\}$. Based on $\leq$, we can define the corresponding strict order $<$, and from we can define the strict relations on subsets of W, $<_{\forall\exists}$ and $<_{\forall\exists 2}$, that coincide on the singletons by axiom (ABS). By the transitivity axioms of $\mathsf{LAP}_{\forall\exists}$ and $\mathsf{LAP}_{\forall\exists 2}$, $\leq$ is a $\odot$-preorder. We define now the evaluation function e, where for each $w \in \Omega$ and $p \in Var$, $e(w, p) = w(p)$. What remains to be checked is that $e_{\mathcal{M}}(\Psi) = v(\Psi)$ for every LAP^1-formula Ψ. In order to prove this equality it suffices to show that, for every $\varphi, \psi \in \mathcal{L}_0$ and $a \in V \setminus \{0\}$, we have $e_{\mathcal{M}}(\varphi \lhd^a \psi) = v(\varphi \lhd^a \psi)$. As mentioned above the proof is very similar to the one in Theorem 3 for all the atomic preference formulas, but specially when $\lhd \in \{\preceq_\alpha, \preceq_\beta\}$. Therefore we only prove the equality for atomic preference formulas of type $\varphi \prec^a_\beta \psi$. By the semantics of LAP^1,

$$e_{\mathcal{M}}(\varphi \prec^a_\beta \psi) = 1 \quad \text{iff} \quad \inf_{\omega' \in [\psi]_{\mathcal{M}}} \sup_{\omega \in [\varphi]_{\mathcal{M}}} [\omega < \omega'] \geq a.$$

By axioms (BS5) and (BS6), we have that LAP^1 proves

$$\varphi \prec^a_\beta \psi \ \leftrightarrow \bigwedge_{\omega' \in \Omega : \omega'(\psi) = 1} \ \bigvee_{\omega \in \Omega : \omega(\varphi) = 1} \overline{\omega} \preceq^a \overline{\omega'}.$$

Therefore, $v(\varphi \prec^a_\beta \psi) = 1$ iff for all $\omega' \in \Omega$ such that $\omega'(\psi) = 1$, there exists $\omega \in \Omega$ such that $\omega(\varphi) = 1$ and $v(\overline{\omega} \prec^a_\beta \overline{\omega'}) = 1$. But $v(\overline{\omega} \prec^a_\beta \overline{\omega'}) = 1$ holds iff $[\omega < \omega'] \geq a$.

In other words, we actually have $v(\varphi \prec^a_\beta \psi) = 1$ iff $\min_{\omega' \in [\psi]_\mathcal{M}} \max_{\omega \in [\varphi]_\mathcal{M}} [\omega \leq \omega'] \geq a$. This concludes the proof. $\square$

In the logic LAP^1 we can define the following modal operators for the indifference relations corresponding to the preference modalities $\preceq^a_\alpha$ and $\preceq^a_\alpha$:

- $\varphi \equiv^a_\alpha \psi$ as $(\varphi \preceq^a_\alpha \psi) \wedge (\psi \preceq^a_\alpha \varphi)$,
- $\varphi \equiv^a_\beta \psi$ as $(\varphi \preceq^a_\beta \psi) \wedge (\psi \preceq^a_\beta \varphi)$,

and, from them, we can in turn define the modalities

- $\varphi \preceq^a_1 \psi$ as $(\varphi \preceq^a_\alpha \psi) \wedge (\varphi \preceq^a_\beta \psi)$,
- $\varphi \equiv^a_1 \psi$ as $(\varphi \equiv^a_\alpha \psi) \wedge (\varphi \equiv^a_\beta \psi)$,
- $\varphi \prec^a_1 \psi$ as $((\varphi \equiv^a_\alpha \psi) \wedge (\varphi \prec^a_\beta \psi)) \vee$
  $((\varphi \equiv^a_\beta \psi) \wedge (\varphi \prec^a_\alpha \psi)) \vee ((\varphi \prec^a_\alpha \psi) \wedge (\varphi \prec^a_\beta \psi))$.

that eventually determine $\langle \preceq_1, \equiv_1, \prec_1 \rangle$ as the preference structure of the logic LAP^1.

We finish this section with one remark justifying the above definition of $\prec^a_1$. Observe that given a preorder $\leq$ on W, the preorder $\preceq_1$ on $\mathcal{P}(\mathrm{W})$ satisfies the following equation:

$$[A \preceq_1 B] = \min([A \preceq_{\forall\exists} B], [A \preceq_{\forall\exists 2} B]),$$

that, by Eq. (1), is equal to

$$\min(\max([A \equiv_{\forall\exists} B], [A <_{\forall\exists} B]), \max([A \equiv_{\forall\exists 2} B], [A <_{\forall\exists 2} B])),$$

and hence, also equal to

$$\begin{aligned} \max(\ & \min([A \equiv_{\forall\exists} B], [A \equiv_{\forall\exists 2} B]), \min([A \equiv_{\forall\exists} B], [A <_{\forall\exists 2} B]), \\ & \min([A <_{\forall\exists} B], [A \equiv_{\forall\exists 2} B]), \min([A <_{\forall\exists} B], [A <_{\forall\exists 2} B])\ ). \end{aligned}$$

Thus, once we define $[A \equiv_1 B] = \min([A \equiv_{\forall\exists} B], [A \equiv_{\forall\exists 2} B])i$, then, again according to Eq. (1), it seems very reasonable to define the strict order value $[A <_1 B]$ by the maximum of the three remaining terms above, that is:

$$\begin{aligned} [A <_1 B] = \max(\ & \min([A \equiv_{\forall\exists} B], [A <_{\forall\exists 2} B]), \\ & \min([A <_{\forall\exists} B], [A \equiv_{\forall\exists 2} B]), \min([A <_{\forall\exists} B], [A <_{\forall\exists 2} B])\ ). \end{aligned}$$

This motivates the definition of $\varphi \prec^a_1 \psi$ above.

5.4 *The Logic* LAP^2

In this subsection we define and study the logic corresponding to the fuzzy preorder $\preceq_2 = \preceq_{\forall\forall}$.

The logic LAP^2 is defined as the expansion of $\mathsf{LAP}_{\forall\forall}$ with modal operators for the strict preference $\prec^a$, for each $a \in V \setminus \{0\}$. We just need to take into account that the semantics for the $\prec^a$ operators is as expected: given a Kripke model $\mathcal{M} = (\mathrm{W}, \leq, e)$,

- $e_{\mathcal{M}}(\varphi \prec^a \psi) = 1$ if $[[\varphi]_{\mathcal{M}} <_{\forall\forall} [\psi]_{\mathcal{M}}] = (\inf_{u \in [\varphi]_{\mathcal{M}}} \inf_{w \in [\psi]_{\mathcal{M}}} [u < w]) \geq a.$

Definition 7 The axioms for LAP^2 are the ones of $\mathsf{LAP}_{\forall\forall}$ for the $\preceq^a$ operators plus:

(AS1) $(\varphi \prec^a \psi) \to (\varphi' \prec^a \psi')$, where $\varphi' \to \varphi, \psi' \to \psi$ are tautologies of CPC
(AS2) $\neg(\varphi \prec^a \varphi)$ (irreflexivity)
(AS3) $(\varphi \prec^a \psi) \wedge (\psi \prec^b \chi) \to (\varphi \prec^{a \odot b} \chi)$ ($\odot$-transitivity)
(AS4) $(\varphi \prec^a \psi) \to (\varphi \prec^b \psi)$, for all $a \leq b$ (nestedness)
(AS5) $(\varphi \vee \psi \prec^a \chi) \leftrightarrow (\varphi \prec^a \chi) \wedge (\psi \prec^a \chi)$ (Left-OR)
(AS6) $(\psi \prec^a \varphi \vee \chi) \leftrightarrow (\psi \prec^a \chi) \wedge (\psi \prec^a \chi)$ (Right-OR)
(SA1) $\bigwedge \left((\overline{\omega} \preceq^a \overline{\omega'}) \to (\overline{\omega'} \preceq^a \overline{\omega}) : a > 0 \right) \to \neg(\overline{\omega} \prec^{a_0} \overline{\omega'})$, where a_0 is the minimum element of $V \setminus \{0\}$.
(SA2) $\neg \bigwedge \left((\overline{\omega} \preceq^a \overline{\omega'}) \to (\overline{\omega'} \preceq^a \overline{\omega}) : a > 0 \right) \to \left((\overline{\omega} \prec^b \overline{\omega'}) \leftrightarrow (\overline{\omega} \preceq^b \overline{\omega'}) \right)$

The only rule of LAP^2 is modus ponens.

Note that axioms (SA1) and (SA2) above are analogous to the ones in LAP^1, and the remark after the definition LAP^1 justifying them applies here as well.

The completeness theorem is ready and the proof is also analogous to previous ones, thus we omit it.

Theorem 7 *For any set $T \cup \{\Phi\}$ of $\mathcal{L}_1$-formulas, it holds that $T \models_{LAP^1} \Phi$ if, and only if, $T \vdash_{LAP^1} \Phi$.*

Finally, let us observe that in LAP^2 we can also define now the preference structure $\langle \preceq_2, \equiv_2, \prec_2 \rangle$ in the obvious way:

- The weak preference statement $\varphi \preceq_2 \psi$ is defined as $\varphi \preceq^a \psi$,
- The equivalence statement $\varphi \equiv_2 \psi$ is defined as $(\varphi \preceq^a \psi) \wedge (\psi \preceq^a \varphi)$,
- The strict preference statement $\varphi \prec_2 \psi$ is defined by $(\varphi \prec^a \psi)$.

Notice however that, strictly speaking, $\varphi \preceq_2 \psi$ is not a fuzzy preorder and $\equiv_2$ is not a fuzzy similarity since they are not reflexive.

6 Conclusions and Future Work

In this paper we have studied preference structures on classical sets arising from fuzzy preference relations, a topic that, as far as we know, has not been very studied in the literature. We have approached the question both from a relational and logical points of view. In the relational approach we have studied and characterized possible extensions of fuzzy preorders on a crisp set W (interpreted as fuzzy preferences between the elements of W) to crisp subsets of W (fuzzy preferences on crisps subsets). Within the logical approach, we have defined and studied several two-tiered modal logics capturing, at the syntactical level, the corresponding preference structures. The same scheme can be generalized to fuzzy preference relations on fuzzy sets. Given a fuzzy preorder $\leq$ on a universe W, we can define corresponding extensions to fuzzy relations on the set $\mathcal{F}(\mathrm{W})$ of fuzzy subsets of W. For example, for all $A, B \in \mathcal{F}(\mathrm{W})$, corresponding extensions for $\forall\exists$ and $\forall\forall$ could be defined as

$$(A \leq_{\forall\exists} B) = [\inf_{u \in W}((\mu_A(u)) \to (\sup_{v \in W}([u \leq v] \odot \mu_B(v))))]$$
$$(A \leq_{\forall\forall} B) = [\inf_{u \in W}((\mu_A(u)) \to (\inf_{v \in W}([u \leq v] \to \mu_B(v))))].$$

As future work we plan to study and characterize these type of fuzzy preference relations taking into account the works by Bodenhofer et al. [2–4], where the authors study some of these relations in the purely fuzzy relational setting. Finally we plan to connect the corresponding fuzzy preference structures with a modal many-valued logic framework, with necessity, possibility, universal and existential modal operators (see [21] for a first approach) in a similar way that it is done in [1] in the classical setting.

Acknowledgements Esteva and Godo acknowledges partial support by the Spanish FEDER /MINECO project TIN2015-71799-C2-1-P. Vidal acknowledges partial support by the joint project of the Austrian Science Fund (FWF) I1897-N25 and the Czech Science Foundation (GACR) 15-34650L, and by the institutional grant RVO:67985807.

References

1. van Benthem, J., Girard, P., Roy, O.: Everything else being equal: a modal logic for ceteris paribus preferences. J. Philos. Logic **38**, 83–125 (2009)
2. Bodenhofer, U.: Orderings of fuzzy sets based on fuzzy orderings part I: the basic approach. Mathw. Soft Comput. **15**(2), 201–218 (2008)
3. Bodenhofer, U.: Orderings of fuzzy sets based on fuzzy orderings part II: generalizations. Mathw. Soft Comput. **15**(3), 219–249 (2008)
4. Bodenhofer, U., Demirci, M.: Strict fuzzy orderings with a given context of similarity. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **16**(2), 147–178 (2008)
5. Domshlak, C., Hüllermeier, E., Kaci, S., Prade, H.: Preferences in AI: an overview. Artif. Intell. **175**, 1037–1052 (2011)
6. Fodor, J., Roubens, M.: Fuzzy Preference Modeling and Multicriteria Decision Support. Kluwer Academic Publishers (1994)
7. Díaz, S., De Baets, B., Montes, S.: Additive decomposition of fuzzy pre-orders. Fuzzy Sets Syst. **158**(8), 830–842 (2007)
8. Díaz, S., De Baets, B., Montes, S.: General results on the decomposition of transitive fuzzy relations. Fuzzy Optim. Decis. Making **9**(1), 1–29 (2010)
9. Díaz, S., Montes, S., De Baets, B.: Transitive decomposition of fuzzy preference relations: the case of nilpotent minimum. Kybernetika **40**(1), 71–88 (2004)
10. Díaz, S., De Baets, B., Montes, S.: On the compositional characterization of complete fuzzy pre-orders. Fuzzy Sets Syst. **159**(17), 2221–2239 (2008)

11. Díaz, S., García-Lapresta, J.L., Montes, S.: Consistent models of transitivity for reciprocal preferences on a finite ordinal scale. Inf. Sci. **178**(13), 2832–2848 (2008)
12. Esteva, F., Godo, L., Rodriguez, R.O., Vetterlein, T.: Logics for approximate and strong entailments. Fuzzy Sets Syst. **197**, 59–70 (2012)
13. Esteva, F., Godo, L., Vidal, A.: On a graded modal logic approach to reason with fuzzy preferences. In: Proceedings of CCIA 2017 (to appear)
14. Fono, L.A., Andjiga, N.G.: Fuzzy strict preference and social choice. Fuzzy Sets Syst. **155**, 372–389 (2005)
15. Halpern, J.Y.: Defining relative likelihood in partially-ordered preference structures. J. Artif. Intell. Res. **7**, 1–24 (1997)
16. Liu, F.: Von Wright's "The Logic of Preference" revisited. Synthese **175**(1), 69–88 (2010)
17. Hansson, S.O., Grüne-Yanoff, T.: Preferences. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy (Winter 2012 Edition). https://plato.stanford.edu/archives/win2012/entries/preferences/ (2012)
18. Kaci, S.: Working with Preferences: Less is More. Springer (2011)
19. Mayor, G., Torrens, J.: On a class of operators for expert systems. Int. J. Intell. Syst. **8**, 771–778 (1993)
20. Sören, H.: On the logic of better. Library of Theoria, no. 2, Uppsala (1957)
21. Vidal, A., Esteva, F., Godo, L.: On finite-valued bimodal logics with an application to reasoning about preferences In: Proceedings of EUSFLAT 2017 (to appear)
22. von Wright, G.H.: The Logic of Preference. Edinburgh University Press (1963)
23. von Wright, G.H.: The logic of preference reconsidered. Theory Decis. **3**(2), 140–169 (1972)

On Possibilistic Dependencies: A Short Survey of Recent Developments

Robert Fullér and István Á. Harmati

Abstract Carlsson and Fullér introduced the notions of possibilistic mean value and variance of fuzzy numbers. Fullér and Majlender introduced a measure of possibilistic covariance between marginal distributions of a joint possibility distribution as the average value of the interactivity relation between the level sets of its marginal distributions. Fullér et al. introduced the possibilistic correlation ratio, the possibilistic correlation coefficient and the possibilistic informational coefficient of correlation. In this paper we give a short survey of some later works which extend and develop these notions.

1 Introduction

In probability theory the notion of mean value of functions of random variables plays a fundamental role in defining the basic characteristic measures of probability distributions: the measure of covariance, variance and correlation of random variables can all be computed as probabilistic means of their appropriately chosen real-valued functions. Similarly, in possibility theory we can use the principle of *average value* of appropriately chosen real-valued functions to define mean value, variance, covariance and correlation of possibility distributions. Marginal probability distributions are determined from the joint one by the principle of 'falling integrals' and marginal possibility distributions are determined from the joint possibility distribution by the principle of 'falling shadows'. Probability distributions can be interpreted as carriers of *incomplete information* [43], and possibility distributions can be interpreted as carriers of *imprecise information*. A function $f\colon [0, 1] \to \mathbb{R}$ is said to be a weighting

R. Fullér (✉)
Department of Informatics, Széchenyi István University, Győr, Hungary
e-mail: rfuller@sze.hu

I.Á. Harmati
Department of Mathematics and Computational Sciences,
Széchenyi István University, Győr, Hungary
e-mail: harmati@sze.hu

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_16

function if f is non-negative, monotone increasing and satisfies the following normalization condition $\int_0^1 f(\gamma)\mathrm{d}\gamma = 1$. Different weighting functions can give different (case-dependent) importances to γ-levels sets of fuzzy numbers. It is motivated in part by the desire to give less importance to the lower levels of fuzzy sets [34] (it is why f should be monotone increasing).

We can define the mean value (variance) of a possibility distribution as the f-weighted average of the probabilistic mean values (variances) of the respective uniform distributions defined on the γ-level sets of that possibility distribution. A measure of possibilistic covariance (correlation) between marginal possibility distributions of a joint possibility distribution can be defined as the f-weighted average of probabilistic covariances (correlations) between marginal probability distributions whose joint probability distribution is defined to be uniform on the γ-level sets of their joint possibility distribution [29]. We should note here that the choice of uniform probability distribution on the level sets of possibility distributions is not without reason. Namely, these possibility distributions are used to represent imprecise human judgments and they carry non-statistical uncertainties. Therefore we will suppose that each point of a given level set is equally possible. Then we apply Laplace's principle of Insufficient Reason: if elementary events are equally possible, they should be equally probable (for more details and generalization of principle of Insufficient Reason see [26], page 59). The main new idea here is to equip the level sets of joint possibility distributions with uniform probability distributions and to derive possibilistic mean value, variance, covariance and correlation of possibility distributions, in such a way that they would be consistent with the extension principle. The idea of equipping the level sets of fuzzy numbers with a uniform probability refers to early ideas of simulation of fuzzy sets by Yager [60], and possibility/probability transforms by Dubois et al. [25] as well as the pignistic transform of Smets [55].

2 Possibilistic Mean Value, Variance, Covariance, Correlation Coefficient and Correlation Ratio

In this section we will recall the possibilistic mean value, variance, covariance and correlation of fuzzy numbers, which are consistent with the extension principle and with the well-known definitions of expectation and variance in probability theory. A *fuzzy number* A is a fuzzy set $\mathbb{R}$ with a normal, fuzzy convex and continuous membership function of bounded support. The family of fuzzy numbers is denoted by $\mathscr{F}$. Fuzzy numbers can be considered as possibility distributions [24, 63]. A fuzzy set C in $\mathbb{R}^2$ is said to be a joint possibility distribution of fuzzy numbers $A, B \in \mathscr{F}$, if it satisfies the relationships $\max\{x \mid C(x, y)\} = B(y)$ and $\max\{y \mid C(x, y)\} = A(x)$ for all $x, y \in \mathbb{R}$. Furthermore, A and B are called the marginal possibility distributions of C.

The possibilistic mean (or expected value), variance, covariance and correlation were originally defined from the measure of possibilistic interactivity (as shown in [10, 29]) but for simplicity, we will present the concept of possibilistic mean value, variance, covariance and possibilistic correlation in a probabilistic setting and point out the fundamental difference between the standard probabilistic approach and the possibilistic one. Let $A \in \mathscr{F}$ be fuzzy number with $[A]^\gamma = [a_1(\gamma), a_2(\gamma)]$ and let U_γ denote a uniform probability distribution on $[A]^\gamma$, $\gamma \in [0, 1]$. Recall that the probabilistic mean value of U_γ is equal to

$$M(U_\gamma) = \frac{a_1(\gamma) + a_2(\gamma)}{2},$$

and its probabilistic variance is computed by

$$\mathrm{var}(U_\gamma) = \frac{(a_2(\gamma) - a_1(\gamma))^2}{12}.$$

In 1987 Dubois and Prade [23] defined an interval-valued expectation of fuzzy numbers, viewing them as consonant random sets. They also showed that this expectation remains additive in the sense of addition of fuzzy numbers. In 2003 Fullér and Majlender [28] introduced the f-weighted *possibilistic mean value* of $A \in \mathscr{F}$ as

$$E_f(A) = \int_0^1 M(U_\gamma)f(\gamma)\mathrm{d}\gamma = \int_0^1 \frac{a_1(\gamma) + a_2(\gamma)}{2} f(\gamma)\mathrm{d}\gamma,$$

where U_γ is a uniform probability distribution on $[A]^\gamma$ for all $\gamma \in [0, 1]$. If $f(\gamma) = 1$ for all $\gamma \in [0, 1]$ then we get

$$E(A) = \int_0^1 M(U_\gamma)f(\gamma)\mathrm{d}\gamma = \int_0^1 \frac{a_1(\gamma) + a_2(\gamma)}{2} \mathrm{d}\gamma,$$

which the possibilistic mean value of A originally introduced in 2001 by Carlsson and Fullér [3]. In 2003 Fullér and Majlender [28] introduced the f-weighted *possibilistic variance* of $A \in \mathscr{F}$ as

$$\mathrm{Var}_f(A) = \int_0^1 \mathrm{var}(U_\gamma)f(\gamma)\mathrm{d}\gamma = \int_0^1 \frac{(a_2(\gamma) - a_1(\gamma))^2}{12} f(\gamma)\mathrm{d}\gamma.$$

In 2004 Fullér and Majlender [29] introduced a measure of possibilistic covariance between marginal distributions of a joint possibility distribution C as the expected value of the interactivity relation between the γ-level sets of its marginal distributions. In 2005 Carlsson et al. [10] showed that the possibilistic covariance between fuzzy numbers A and B can be written as the weighted average of the probabilistic covariances between random variables with uniform joint distribution on the level sets of their joint possibility distribution C. The *f-weighted measure of possibilistic covariance* between $A, B \in \mathcal{F}$ (with respect to their joint distribution C) [29] can be written as,

$$\mathrm{Cov}_f(A, B) = \int_0^1 \mathrm{cov}(X_\gamma, Y_\gamma) f(\gamma) \mathrm{d}\gamma,$$

and the f-weighted *possibilistic correlation coefficient* of $A, B \in \mathcal{F}$ (with respect to their joint distribution C) is defined by [31],

$$\rho_f(A, B) = \int_0^1 \rho(X_\gamma, Y_\gamma) f(\gamma) \mathrm{d}\gamma$$

where

$$\rho(X_\gamma, Y_\gamma) = \frac{\mathrm{cov}(X_\gamma, Y_\gamma)}{\sqrt{\mathrm{var}(X_\gamma)}\sqrt{\mathrm{var}(Y_\gamma)}}$$

and X_γ and Y_γ are random variables whose joint distribution is uniform on $[C]^\gamma$, $\mathrm{cov}(X_\gamma, Y_\gamma)$ denotes their covariance, for all $\gamma \in [0, 1]$.

In statistics, the correlation ratio is a measure of the relationship between the statistical dispersion within individual categories and the dispersion across the whole population or sample. The correlation ratio was originally introduced by Pearson [52] and it was extended to random variables by Kolmogorov [44] as,

$$\eta^2(X|Y) = \frac{\mathrm{var}[M(X|Y)]}{\mathrm{var}(X)},$$

where X and Y are random variables. If X and Y have a joint probability density function, denoted by $g(x, y)$, then we can compute $\eta^2(X|Y)$ using the following formulas

$$M(X|Y = y) = \int_{-\infty}^{\infty} x g(x|y) \mathrm{d}x$$

and

$$\mathrm{var}[M(X|Y)] = M(M(X|y) - M(X))^2,$$

and where,

$$g(x|y) = \frac{g(x, y)}{g(y)}.$$

In 2010 Fullér et al. [30] introduced the *f-weighted possibilistic correlation ratio* of marginal possibility distribution A with respect to marginal possibility distribution B as

$$\eta_f^2(A|B) = \int_0^1 \eta^2(X_\gamma|Y_\gamma) f(\gamma) \mathrm{d}\gamma$$

where X_γ and Y_γ are random variables whose joint distribution is uniform on $[C]^\gamma$, $\text{cov}(X_\gamma, Y_\gamma)$ denotes their covariance and $\eta(X_\gamma|Y_\gamma)$ denotes their probabilistic correlation ratio [44], for all $\gamma \in [0, 1]$.

In 2012 Fullér et al. [32] introduced the *f-weighted possibilistic informational coefficient of correlation*. For any two continuous random variables X and Y (admitting a joint probability density), their mutual information is given by,

$$I(X, Y) = \iint g(x, y) \ln \frac{g(x, y)}{g_1(x)g_2(y)} \mathrm{d}x\, \mathrm{d}y$$

where $g(x, y)$ is the joint probability density function of X and Y, and $g_1(x)$ and $g_2(y)$ are the marginal density functions of X and Y, respectively. The informational coefficient of correlation of X and Y is defined by

$$L(X, Y) = \sqrt{1 - e^{-2I(X,Y)}}$$

Then the *f-weighted possibilistic informational coefficient of correlation* of marginal possibility distributions A and B is defined by

$$L(A, B) = \int_0^1 L(X_\gamma, Y_\gamma) f(\gamma) \mathrm{d}\gamma$$

where X_γ and Y_γ are random variables whose joint distribution is uniform on $[C]^\gamma$, for all $\gamma \in [0, 1]$.

Note 1 There exist several other ways to define correlation coefficient for fuzzy numbers, e.g. Liu and Kao [47] used fuzzy measures to define a fuzzy correlation coefficient of fuzzy numbers and they formulated a pair of nonlinear programs to find the α-cut of this fuzzy correlation coefficient, then, in a special case, Hong [39] showed an exact calculation formula for this fuzzy correlation coefficient. Vaidyanathan [57] introduced a new measure for the correlation coefficient between triangular fuzzy variables called credibilistic correlation coefficient.

Fullér and colleagues have extensively used the possibilistic mean value, variance, covariance and correlation in their later works for real option valuation [4, 8, 14], portfolio selection problems [6, 7, 11, 12] and strategic planning [9, 13]. For example, in 2007 Carlsson et al. [12] developed a methodology for valuing options on R&D projects, when future cash flows are estimated by trapezoidal fuzzy numbers. In particular, they presented the following fuzzy mixed integer programming model for the R&D optimal portfolio selection problem,

$$\begin{aligned}&\text{maximize} \sum_{i=1}^{N} u_i \mathscr{F}_i \\ &\text{subject to} \sum_{i=1}^{N} u_i X_i + \sum_{i=1}^{N} (1-u_i) c_i \leq B \\ &u_i \in \{0,1\},\ i = 1, \ldots N,\end{aligned}$$

where N is the number of R&D projects, B is the whole investment budget, u_i is the decision variable that takes value one if the i-th project should start now (at time zero) or takes the value zero if it should be postponed and started at a later time, c_i denotes the cost of the postponment (i.e. keep the option alive), X_i is the investment cost, and $\mathscr{F}_i$ denotes the possibilistic deferral flexibility of the i-th project for $i = 1, \ldots, N$. Furthermore, they discussed how their methodology can be used to build decision support tools for optimal R&D project selection in a corporate environment. They also claimed that the imprecision we encounter when judging or estimating future cash flows is not stochastic in nature, and the use of probability theory gives us a misleading level of precision and a notion that consequences somehow are repetitive. This is not the case, the uncertainty is genuine, i.e. we simply do not know the exact levels of future cash flows. Without introducing fuzzy real option models it would not be possible to formulate this genuine uncertainty.

In 2009 Collan et al. [18] presented a new method (fuzzy pay-off method) for real option valuation using fuzzy numbers that is based on findings from earlier real option valuation methods and from fuzzy real option valuation. They also presented the use of number of different types of fuzzy numbers with the method and an application of the new method in an industry setting. In 2010 Carlsson et al. [13] used fuzzy real option models for the problem of closing/not closing a production plant in the forest products industry sector. In 2013 Carlsson and Fullér [15] implemented a hybrid probabilistic and possibilistic model to assess the success of computing tasks in a Grid. Using the predictive probabilistic approach they developed a framework for resource management in grid computing, and by introducing an upper limit for the number of possible failures, they approximated the probability that a particular computing task can be executed. Coroianu and Fullér [19] studied he problem of additivity property of the weighted possibilistic mean operator for interactive fuzzy numbers. They showed that the weighted possibilistic mean operator is additive on the set of symmetric fuzzy numbers if their joint possibility distribution is defined by a triangular norm. They also showed some results for general joint-possibility-distribution-based additions of fuzzy numbers of symmetrical opposite sides.

3 Recent Developments

The notions of possibilistic mean value, variance, covariance and correlation are used in many different research areas and by many different authors (Google Scholar finds over 2,000 citations to papers [3, 8, 10, 12, 28, 29]).

In 2005 Yoshida et al. [61] evaluated the randomness and fuzziness in fuzzy stochastic processes by the probabilistic expectation and the mean values defined by fuzzy measures and λ-weighting functions. The mean values are demonstrated particularly in three kinds of important fuzzy measures: possibility measure, necessity measure and credibility measure. Furthermore, by introducing fuzziness to stochastic processes in optimization/decision-making, they considered a new model with uncertainty of both randomness and fuzziness, which is a reasonable and natural extension of the original stochastic process.

In 2006 Fang et al. [27] proposed a portfolio rebalancing model with transaction costs based on fuzzy decision theory and illustrated the behaviour of their proposed model using real data from the Shanghai Stock Exchange. Huang [41] selected the optimal portfolio with fuzzy returns by criteria of chance represented by credibility measure. He introduced two types of credibility-based portfolio selection models: (i) by one chance criterion, the objective is to maximize the investor's return at a given threshold confidence level; (ii) by another chance criterion, the objective is to maximize the credibility of achieving a specified return level subject to the constraints. To solve the resulting problems he designed a hybrid intelligent algorithm integrating fuzzy simulation and genetic algorithm.

In 2007 Zhang et al. [64] proposed two kinds of portfolio selection models based on lower and upper possibilistic means and possibilistic variances, respectively, and introduced the notions of lower and upper possibilistic efficient portfolios. They also presented an algorithm which can derive the explicit expression of the possibilistic efficient frontier for the possibilistic mean-variance portfolio selection problem dealing with lower bounds on asset holdings. Zhang and Wang [65] investigated the relationship between several crisp possibilistic variances and covariances of fuzzy numbers. Silva et al. [54] presented and developed an original and novel fuzzy sets based method that solves a class of quadratic programming problems with vagueness in the set of constraints. The method uses two phases to solve fuzzy quadratic programming problems, which eventually can be considered in the portfolio context. In the first phase they parametrize the fuzzy problem in several classical alpha-problems with different cutting levels. In the second phase each of these alpha-problems is solved by using conventional solving techniques.

In 2008 by introducing the concept of semivariance of fuzzy variable Huang [42] proposed two fuzzy mean-semivariance models for portfolio selection problems in fuzzy environment. To solve the new models in general cases, he presented a fuzzy simulation based genetic algorithm. By morphing mean-variance optimization portfolio model into semi-absolute deviation model, Gupta et al. [36] applied multi criteria decision making via fuzzy mathematical programming to develop comprehensive models of asset portfolio optimization for the investors' pursuing either of the aggressive or conservative strategies.

In 2009 Chen et al. [16] considered a possibilistic mean-variance (FMVC) portfolio selection model and proposed a cutting plane algorithm to solve it. Xu et al. [59] presented a fuzzy normal jump-diffusion model for European option pricing, with uncertainty of both randomness and fuzziness in the jumps, which is a reasonable and a natural extension of the Merton [50] normal jump-diffusion model. Based on the crisp weighted possibilistic mean values of the fuzzy variables in fuzzy normal jump-diffusion model, they also obtained the crisp weighted possibilistic mean normal jump-diffusion model. Yoshida [62] discussed value-at-risk portfolio model under uncertainty. In his proposed model the means, the variances and the measurements of imprecision for fuzzy numbers/fuzzy random variables are evaluated in the possibility case and the necessity case, and the rate of return in portfolio is estimated regarding the both random factors and imprecise factors. Zhang et al. [66] proposed a new portfolio selection model with the maximum utility based on the interval-valued possibilistic mean and possibilistic variance, which is a two-parameter quadratic programming problem. They also presented a sequential minimal optimization (SMO) algorithm to obtain the optimal portfolio. The remarkable feature of their algorithm is that it is extremely easy to implement, and it can be extended to any size of portfolio selection problems for finding an exact optimal solution.

In 2010 Zhang et al. [67] proposed a possibilistic portfolio adjusting model with transaction costs and bounded constraints on holdings of assets, which can be transformed into a linear programming problem. Both the lower bounds on holdings and the total investment constraints influence the optimal portfolio adjusting strategies. Gładysz and Kasperski [33] discussed the problem of computing the mean absolute deviation in a set of uncertain variables. The uncertainty is modelled by closed intervals and fuzzy intervals. Some polynomial algorithms for determining the lower and upper bounds for the mean absolute deviation under interval uncertainty are proposed. Possibility theory is then applied to generalize the interval uncertainty representation to the fuzzy one.

In 2011 Ho and Liao [38] proposed a fuzzy binomial approach for investment project valuation in uncertain environments from the aspect of real options. Their approach also reveals the value of flexibilities embedded in the project. Duan and Stahlecker [22] considered static portfolio selection problem, in which future returns of securities are given as fuzzy sets. In contrast to traditional analysis, they assume that investment decisions are not based on statistical expectation values, but rather on maximal and minimal potential returns resulting from the α-cuts of these fuzzy sets. By aggregating over all α-cuts and assigning weights for both best and worst possible cases they get a new objective function to derive an optimal portfolio. Lee and Lee [45] examined the strategic characteristic of RFID (Radio Frequency Identification) investment and proposed a fuzzy real options technique that can consider various situations of expected cash flow or investment costs as a plan to support investment decisions.

In 2012 Deng and Li [20] proposed a portfolio selection model with borrowing constraint by means of possibilistic mean, possibilistic variance, and possibilistic covariance under the assumption that the returns of assets are fuzzy numbers. They presented a quadratic programming model with inequality constraints when the returns of assets are trapezoid fuzzy numbers and utilized the Lemke algorithm to solve the problem. Zhang et al. [68] dealt with a multi-period portfolio selection problem with fuzzy returns and presented a possibilistic mean-semivariance-entropy model for multi-period portfolio selection by taking into account four criteria: return, risk, transaction cost and diversification degree of portfolio. In their proposed model, the return level is quantified by the possibilistic mean value of return, the risk level is characterized by the lower possibilistic semivariance of return, and the diversification degree of portfolio is measured by the originally presented possibilistic entropy. Based on the possibilistic mean and the possibilistic variance/covariance of fuzzy numbers, Chrysafis [17] proposed a method to reduce some problems arising from the Capital Asset Pricing Model (CAPM) assumptions.

In 2013 Thavaneswaran et al. [56] used fuzzy set theory to price binary options. Namely, they studied binary options by fuzzifying the maturity value of the stock price using trapezoidal, parabolic and adaptive fuzzy numbers. Hsieh and Tsaur [40] proposed a simplified fuzzy regression equation based on possibilistic mean and variance method and used it for modeling the constraints and objective function of a fuzzy regression model without determining the membership function of extrapolative values. Liu and Zhang [48] discussed a multi-objective portfolio optimization problem for practical portfolio selection in fuzzy environment, in which the return rates and the turnover rates are characterized by fuzzy variables. Based on the possibility theory, they quantified fuzzy return and liquidity by possibilistic mean, and market risk and liquidity risk are measured by lower possibilistic semivariance. They proposed a fuzzy multi-objective programming technique to transform their proposed models into corresponding single-objective models and then designed a genetic algorithm for their solution.

In 2014 Wang et al. [58] employed the weighted possibilistic mean (WPM), weighted interval valued possibilistic mean (WIVPM) of fuzzy number as a sort of representative values for the fuzzy attribute data, and establish new fuzzy control charts with WPM and WIVPM. They compared the performance of the charts to the existing fuzzy charts with a fuzzy *c*-chart example via newly defined average number of inspection for variation of control state. Based on possibility theory and the assumption that the returns of assets are triangular fuzzy numbers, Deng and Li [21] proposed a bi-objective nonlinear portfolio selection model. They show that their nonlinear bi-objective model is equivalent to the linear bi-objective minimizing programming model on the basis of possibilistic mean and possibilistic variance.

In 2015 Nguyen et al. [51] initiated the fuzzy Sharpe ratio in the fuzzy modeling context. In addition to the introduction of the new risk measure, they also put forward the reward-to-uncertainty ratio to assess the portfolio performance in fuzzy modeling. Zhang [69] considered a multi-period portfolio selection problem in a fuzzy investment environment, in which the return and risk of assets are characterized by possibilistic mean value and possibilistic semivariance, respectively. Based on the theories of possibility, he proposed a new multi-period possibilistic portfolio selection model, which contains risk control, transaction costs, borrowing constraints, threshold constraints and cardinality constraints. By redefining the concepts of mean and variance for fuzzy numbers, Li et al. [46] formulated a fuzzy mean-variance-skewness portfolio selection model.

In 2016 Mashayekhi and Omrani [49] proposed a novel multi objective model for portfolio selection, where the asset returns are considered as trapezoidal fuzzy numbers. Their model incorporates the DEA cross-efficiency into Markowitz mean-variance model and considers return, risk and efficiency of the portfolio. Rubio et al.

[53] proposed the weighted fuzzy time series method to forecast the future performance of returns on portfolios. They modelled the uncertain parameters of the fuzzy portfolio selection models using a possibilistic interval-valued mean approach, and approximate the uncertain future return on a given portfolio by means of a trapezoidal fuzzy number. Guo et al. [35] considered a fuzzy multi-period portfolio selection problem with V-Shaped transaction cost. Compared with the traditional studies assuming that assets have the same investment horizon, they handled the practical but complicated situation in which assets have different investment horizons. Within the framework of credibility theory, they formulate a mean-variance model with the objective of maximizing the terminal return under the total risk constraint over the whole investment.

In 2017 Babazadeh et al. [1] presented a multi-objective possibilistic programming model to design a second-generation biodiesel supply chain network under risk. Their model minimizes the total costs of biodiesel supply chain from feedstock supply centers to customer centers besides minimizing the environmental impact of all involved processes under a well-to-wheel perspective. Brunelli and Mezei [2] presented an analysis of approximate operations on fuzzy numbers. By focusing on the ranking and defuzzification procedures as essential tools in fuzzy decision making problems, they studied the errors produced by the application of approximate operations.

4 Concluding Remarks

Possibility theory is mathematically the simplest uncertainty theory for dealing with incomplete information [26]. This may be the reason why possibilistic dependencies are used in many different research areas like information sciences, geosciences, social sciences, economics, mathematical and computer modelling, financial engineering, systems engineering, military engineering, and robotics. We have shown several applications of possibilistic dependencies ranging from multi-period portfolio selection problem with fuzzy returns to designing a second-generation biodiesel supply chain network. However, it is still an open problem to construct joint possibility distribution for correlated variables in applications [37].

References

1. Babazadeh, R., Razmi, J., Pishvaee, M.S., Rabbani, M.: A sustainable second-generation biodiesel supply chain network design problem under risk. Omega **66**, 258–277 (2017)
2. Brunelli, M., Mezei, J.: An inquiry into approximate operations on fuzzy numbers. Int. J. Approx. Reas. **81**, 147–159 (2017)
3. Carlsson, C., Fullér, R.: On possibilistic mean value and variance of fuzzy numbers. Fuzzy Sets Syst. **122**, 315–326 (2001)

4. Carlsson, C., Fullér, R.: On optimal investment timing with fuzzy real options. In: Proceedings of the EUROFUSE 2001 Workshop on Preference Modelling and Applications, April 25–28, pp. 235–239. Granada, Spain (2001)
5. Carlsson, C., Fullér, R., Majlender, P.: Project selection with fuzzy real options. In: Proceedings of the Second International Symposium of Hungarian Researchers on Computational Intelligence, pp. 81–88 (2001)
6. Carlsson, C., Fullér, R.: Project scheduling with fuzzy real options. In: Trappl, R. (ed.) Cybernetics and Systems 2002, Proceedings of the Sixteenth European Meeting on Cybernetics and Systems Research, Vienna, April 2–4, 2002, Austrian Society for Cybernetic Studies, pp. 511–513 (2002)
7. Carlsson, C., Fullér, R., Majlender, P.: A possibilistic approach to selecting portfolios with highest utility score. Fuzzy Sets Syst. **131**, 13–21 (2002)
8. Carlsson, C., Fullér, R.: A fuzzy approach to real option valuation. Fuzzy Sets Syst. **139**, 297–312 (2003)
9. Carlsson, C., Fedrizzi, M., Fullér, R.: Fuzzy Logic in Management, International Series in Operations Research and Management Science, vol. 66. Kluwer Academic Publishers, Boston (2004)
10. Carlsson, C., Fullér, R., Majlender, P.: On possibilistic correlation. Fuzzy Sets Syst. **155**, 425–445 (2005)
11. Carlsson, C., Fullér, R., Majlender, P.: A fuzzy real options model for R&D project evaluation. In: Liu, Y., Chen, G., Ying, M. (eds.) Proceedings of the Eleventh IFSA World Congress, Beijing, China, July 28–31, 2005, pp. 1650–1654. Tsinghua University Press and Springer, Beijing (2005)
12. Carlsson, C., Fullér, R., Heikkilä, M., Majlender, P.: A fuzzy approach to R&D project portfolio selection. Int. J. Approximate Reasoning **44**, 93–105 (2007)
13. Carlsson, C., Heikkilä, M., Fullér, R.: Fuzzy real options models for closing/not closing a production plant. In: Kahraman, C., Yavuz, M. (eds.) Production Engineering and Management under Fuzziness, Studies in Fuzziness and Soft Computing, vol. 252, pp. 537–560. Springer [ISBN 978-3-642-12051-0] (2010)
14. Carlsson, C., Fullér, R.: Possibility for Decision: A Possibilistic Approach to Real Life Decisions, Studies in Fuzziness and Soft Computing Series, vol. 270. Springer (2011)
15. Carlsson, C., Fullér, R.: Probabilistic versus possibilistic risk assessment models for optimal service level agreements in grid computing. Inf. Syst. e-Bus. Manage. **11**, 13–28 (2013)
16. Chen, G., Liao, X., Wang, S.: A cutting plane algorithm for MV portfolio selection model. Appl. Math. Comput. **215**, 1456–1462 (2009)
17. Chrysafis, K.A.: Corporate investment appraisal with possibilistic CAPM. Math. Comput. Modell. **55**, 1041–1050 (2012)
18. Collan, M., Fullér, R., Mezei, J.: A fuzzy pay-off method for real option valuation. J. Appl. Math. Decis. Sci. **2009**, Article ID 238196 (2009)
19. Coroianu, L., Fullér, R.: On additivity of the weighted possibilistic mean operator. In: Proceedings of the Fourteenth IEEE International Symposium on Computational Intelligence and Informatics, November 19–21, pp. 303–308. Budapest, Hungary (2013)
20. Deng, X., Li, R.: A portfolio selection model with borrowing constraint based on possibility theory. Appl. Soft Comput. **12**, 754–758 (2012)
21. Deng, X., Li, R.: Gradually tolerant constraint method for fuzzy portfolio based on possibility theory. Inf. Sci. **259**, 16–24 (2014)
22. Duan, L., Stahlecker, P.: A portfolio selection model using fuzzy returns. Fuzzy Optim. Decis. Making **10**, 167–191 (2011)
23. Dubois, D., Prade, H.: The mean value of a fuzzy number. Fuzzy Sets Syst. **24**, 279–300 (1987)
24. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
25. Dubois, D., Prade, H., Sandri, S.: On possibility/probability transformations. In: Lowen, R., Roubens, M. (eds.) Fuzzy Logic: State of the Art, pp. 103–112. Dordrecht, Kluwer Academic Publishers (1993)

26. Dubois, D.: Possibility theory and statistical reasoning. Comput. Stat. Data Anal. **51**, 47–69 (2006)
27. Fang, Y., Lai, K.K., Wang, S.-Y.: Portfolio rebalancing model with transaction costs based on fuzzy decision theory. Eur. J. Oper. Res. **175**, 879–893 (2006)
28. Fullér, R., Majlender, P.: On weighted possibilistic mean and variance of fuzzy numbers. Fuzzy Sets Syst. **136**, 363–374 (2003)
29. Fullér, R., Majlender, P.: On interactive fuzzy numbers. Fuzzy Sets Syst. **143**, 355–369 (2004)
30. Fullér, R., Mezei, J., Várlaki, P.: A correlation ratio for possibility distributions. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) Computational Intelligence for Knowledge-Based Systems Design, Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2010), June 28–July 2, 2010, Dortmund, Germany. Lecture Notes in Artificial Intelligence, vol. 6178, pp. 178–187. Springer, Berlin (2010)
31. Fullér, R., Mezei, J., Várlaki, P.: An improved index of interactivity for fuzzy numbers. Fuzzy Sets Syst. **165**, 56–66 (2011)
32. Fullér, R., Harmati, I.Á., Várlaki, P., Rudas, I.: On weighted possibilistic informational coefficient of correlation. Int. J. Math. Models Meth. Appl. Sci. **6**, 592–599 (2012)
33. Gładysz, B., Kasperski, A.: Computing mean absolute deviation under uncertainty. Appl. Soft Comput. **10**, 361–366 (2010)
34. Goetschel, R., Voxman, W.: Elementary fuzzy calculus. Fuzzy Sets Syst. **18**, 31–43 (1986)
35. Guo, S., Yu, L., Li, X., Kar, S.: Fuzzy multi-period portfolio selection with different investment horizons. Eur. J. Oper. Res. **254**, 1026–1035 (2016)
36. Gupta, P., Kumar, M., Mehlawat, A.: Saxena Asset portfolio optimization using fuzzy mathematical programming. Inf. Sci. **178**, 1734–1755 (2008)
37. He, Y., et al.: An uncertainty visualization technique using possibility theory: possibilistic marching cubes. Int. J. Uncertainty Quantification **5**, 433–451 (2015)
38. Ho, S.-H., Liao, S.-H.: A fuzzy real option approach for investment project valuation. Expert Syst. Appl. **38**, 15296–15302 (2011)
39. Hong, D.H.: Fuzzy measures for a correlation coefficient of fuzzy numbers under T_W (the weakest t-norm)-based fuzzy arithmetic operations. Inf. Sci. **176**, 150–160 (2006)
40. Hsieh, W.-Y., Tsaur, R.-C.: Epidemic forecasting with a new fuzzy regression equation. Qual. Quant. **47**, 3411–3422 (2013)
41. Huang, X.: Fuzzy chance-constrained portfolio selection. Appl. Math. Comput. **177**, 500–507 (2006)
42. Huang, X.: Mean-semivariance models for fuzzy portfolio selection. J. Comput. Appl. Math. **217**, 1–8 (2008)
43. Jaynes, E.T.: Probability Theory: The Logic of Science. Cambridge University Press (2003)
44. Kolmogorov, A.N.: Grundbegriffe der Wahrscheinlichkeitsrechnung, p. 62. Julius Springer, Berlin (1933)
45. Lee, Y.-C., Lee, S.-S.: The valuation of RFID investment using fuzzy real option. Expert Syst. Appl. **38**, 12195–12201 (2011)
46. Li, X., Guo, S., Yu, L.: Skewness of fuzzy numbers and its applications in portfolio selection. IEEE Trans. Fuzzy Syst. **23**, 2135–2143 (2015)
47. Liu, S.T., Kao, C.: Fuzzy measures for correlation coefficient of fuzzy numbers. Fuzzy Sets Syst. **128**, 267–275 (2002)
48. Liu, Y.-J., Zhang, W.-G.: Fuzzy portfolio optimization model under real constraints. Insur. Math. Econ. **53**, 704–711 (2013)
49. Mashayekhi, Z., Omrani, H.: An integrated multi-objective Markowitz-DEA cross-efficiency model with fuzzy returns for portfolio selection problem. Appl. Soft Comput. **38**, 1–9 (2016)
50. Merton, R.C.: Option pricing when underlying stock returns are discontinuous. J. Financ. Econ. **3**, 125–144 (1976)
51. Nguyen, T.T., Gordon-Brown, L., Khosravi, A., Creighton, D., Nahavandi, S.: Fuzzy portfolio allocation models through a new risk measure and fuzzy sharpe ratio. IEEE Trans. Fuzzy Syst. **23**, 656–676 (2015)

52. Pearson, K.: On a new method of determining correlation, when one variable is given by alternative and the other by multiple categories. Biometrika **7**(3), 248–257 (1910)
53. Rubio, A., Bermdez, J.D., Vercher, E.: Forecasting portfolio returns using weighted fuzzy time series methods. Int. J. Approx. Reason. **75**, 1–12 (2016)
54. Silva, R.C., Verdegay, J.L., Yamakami, A.: Two-phase method to solve fuzzy quadratic programming problems. In: IEEE International Fuzzy Systems Conference, vol. 2007, pp. 1–6. London (2007)
55. Smets, P.: Constructing the pignistic probability function in a context of uncertainty. In: Henrion, M., et al. (eds.) Uncertainty in Artificial Intelligence, vol. 5, pp. 29–39. North-Holland, Amsterdam (1990)
56. Thavaneswaran, A., Appadoo, S.S., Frank, J.: Binary option pricing using fuzzy numbers. Appl. Math. Lett. **26**, 65–72 (2013)
57. Vaidyanathan, V.S.: Correlation of triangular fuzzy variables using credibility theory. Int. J. Comput. Cogn. **8**, 21–23 (2010)
58. Wang, D., Li, P., Yasuda, M.: Construction of fuzzy control charts based on weighted possibilistic mean. Commun. Stat. Theory Meth. **43**, 3186–3207 (2014)
59. Xu, W., Wu, C., Xu, W., Li, H.: A jump-diffusion model for option pricing under fuzzy environments. Insur. Math. Econ. **44**(200), 337–344
60. Yager, R.R.: A procedure for ordering fuzzy subsets of the unit interval. Inf. Sci. **24**, 143–161 (1981)
61. Yoshida, Y., Yasuda, M., Nakagami, J.-I., Kurano, M.: A discrete-time American put option model with fuzziness of stock prices. Fuzzy Optim. Decis. Making **4**, 191–207 (2005)
62. Yoshida, Y.: An estimation model of value-at-risk portfolio under uncertainty. Fuzzy Sets Syst. **160**, 3250–3262 (2009)
63. Zadeh, L.A.: Concept of a linguistic variable and its application to approximate reasoning, **I, II, III**. Inf. Sci. **8**(1975) 199–249, 301–357; 9(1975) 43–80
64. Zhang, W.-G., Wang, Y.-L., Chen, Z.-P., Nie, Z.-K.: Possibilistic mean-variance models and efficient frontiers for portfolio selection problem. Inf. Sci. **177**, 2787–2801 (2007)
65. Zhang, W.-G., Wang, Y.-L.: A comparative analysis of possibilistic variances and covariances of fuzzy numbers. Fundamenta Informaticae **79**, 257–263 (2007)
66. Zhang, W.-G., Zhang, X.-L., Xiao, W.-L.: Portfolio selection under possibilistic mean-variance utility and a SMO algorithm. Eur. J. Oper. Res. **197**, 693–700 (2009)
67. Zhang, W.-G., Xiao, W.-L., Xu, W.-J.: A possibilistic portfolio adjusting model with new added assets. Econ. Modell. **2**, 208–213 (2010)
68. Zhang, W.-G., Liu, Y.-J., Xu, W.-J.: A possibilistic mean-semivariance-entropy model for multi-period portfolio selection with transaction costs. Eur. J. Oper. Res. **222**, 341–349 (2012)
69. Zhang, P.: Multi-period possibilistic mean semivariance portfolio selection with cardinality constraints and its algorithm. J. Math. Modell. Algorithms Oper. Res. **14**, 239–253 (2015)

Penalty Function in Optimization Problems: A Review of Recent Developments

Humberto Bustince, Javier Fernandez and Pedro Burillo

Abstract In this chapter we make a brief revision of some recent developments on the notion of penalty function as a tool for the fusion of information, including the most recently published definition as well as the extension of the concept to the lattice setting.

1 Introduction

Penalty functions provide a method to determine up to what extent a given output is similar (or dissimilar) to a set of inputs. This information can be used, by means of an appropriate minimization procedure, to define a function (the so-called penalty-based function) for fusing the considered inputs. In this sense, it is a powerful tool for those applications where fusion of information is crucial, as it is the case for almost every real-world application. For this reason, and since it was first suggested by Yager in 1993 [18], the notion of penalty and penalty-based functions have been gaining an increasing interest among the scientific community, specially in order to overcome the different technical difficulties which arise to properly define such a class of functions, see [6] for a complete discussion.

In this chapter, and basing ourselves in [6], we make a brief review of the evolution of the ideas of penalty and penalty-based functions, from its origin in the works of Yager up to the last definition in the literature, which has finally encompassed all the desired properties for such functions. Furthermore, we also discuss briefly how this concept can also be extended to more general setting, considering in particular the case of a Cartesian product of lattices, which is of interest in applications such as image processing [3, 8].

H. Bustince (✉) · J. Fernandez · P. Burillo
Departamento de Automatica y Computacion, Universidad Publica de Navarra,
Campus Arrosadia s/n, 31006 Pamplona, Spain
e-mail: bustince@unavarra.es

J. Fernandez
e-mail: fcojavier.fernandez@unavarra.es

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_17

The structure of this chapter is as follows. First we discuss some preliminary notions and results, and, in Sect. 3, we make the review of the evolution of the idea of penalty functions and penalty-based function. In Sect. 4, we provide an idea of how a similar concept can be defined in a Cartesian product of lattices. We finish with some conclusions and references.

2 Preliminaries

We denote by $\mathbb{I}$ a closed subinterval of the extended real line, i.e., $\mathbb{I} = [a, b] \subseteq \mathbb{R}$.

The notion of penalty-function relies on the ideas of convexity and quasi-convexity, that we recall now.

Definition 1 A function $f : \mathbb{I} \to \mathbb{R}$ is convex if for every $x, y \in \mathbb{I}$ and for every $\lambda \in [0, 1]$ the inequality

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

holds.

Definition 2 A function $f : \mathbb{I} \to \mathbb{R}$ is quasi-convex if for every $x, y \in \mathbb{I}$ and for every $\lambda \in [0, 1]$ the inequality

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}$$

holds.

Quasi-convex functions are relevant for us due to the following result about minimization [15]:

Proposition 1 *Let $f : \mathbb{I} \to \mathbb{R}$ be a quasi-convex function. Then, the set of minimizers of f is either empty or a connected set.*

We also recall here the idea of lower semicontinuity.

Definition 3 A function $f : \mathbb{I} \to \mathbb{R}$ is lower semicontinuous at $x_0 \in \mathbb{I}$ if

$$\liminf_{x \to x_0} f(x) \geq f(x_0).$$

Analogously one defines upper semicontinuity. Observe that a function is continuous at $x_0 \in \mathbb{I}$ if and only if it is upper and lower semicontinuous there.

Again, lower semicontinuity is important for us because, in a compact domain, the set of minimizers of a lower semicontinuous function is not empty.

Proposition 2 *Let $f : \mathbb{I} \to \mathbb{R}$ be a lower semicontinuous function, with $\mathbb{I}$ bounded. Then, the set of minimizers of f is not empty.*

We finally review the notions related to aggregation functions. We take, from now on, $n \geq 2$.

Definition 4 [2, 12] A function $A : [0,1]^n \to [0,1]$ is said to be an n-ary aggregation function if:

(A1) A is increasing in each argument: for each $i \in \{1,\dots,n\}$, if $x_i \leq y$, then $A(x_1,\dots,x_n) \leq A(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_n)$;
(A2) A satisfies the boundary conditions: $A(0,\dots,0) = 0$ and $A(1,\dots,1) = 1$.

It is well known that an aggregation function $f : [0,1]^n \to [0,1]$ is called averaging if, for all $(x_1,\dots,x_n) \in [0,1]^n$, it holds that:

$$\min\{x_1,\dots,x_n\} \leq f(x_1,\dots,x_n) \leq \max\{x_1,\dots,x_n\}.$$

In particular, an aggregation function is averaging if and only if it is idempotent, i.e., $f(x,\dots,x) = x$ for every $x \in [0,1]$.

Finally, and for the sake of completeness, we recall that a fuzzy set A on an universe U is a mapping $A : U \to [0,1]$. The value $A(u)$ is called membership degree of the element u. We will denote by $\mathscr{F}\mathscr{S}(U)$ the class of all fuzzy sets defined on the referential U.

3 The Evolution of the Idea of Penalty Functions

The first approaches to the notion of penalty function in information aggregation procedures may be traced back to Yager [18]. A few years later, in 1997, Yager and Rybalov [19] considered the possibility of obtaining an appropriate function in order to fuse information by means of a minimization procedure. The information to be fused may even be of different nature. Formally, and considering only numerical dates, their proposal was the following:

Definition 5 The function $LP : \mathbb{R}^2 \to \mathbb{R}^+$ is said to be a local penalty function if, for any $x_i, x_j, y \in \mathbb{R}$ and $i, j = 1,\dots,n$, it satisfies:

(LP3.1-1) $LP(x_i, y) = 0$, if $x_i = y$;
(LP3.1-2) $LP(x_i, y) > 0$, if $x_i \neq y$;
(LP3.1-3) $LP(x_i, y) \geq LP(x_j, y)$, if $| x_i - y | > | x_j - y |$,

where y is called the fused value related to each observation in $\mathbf{x}$.

Definition 6 Let $LP : \mathbb{R}^2 \to \mathbb{R}^+$ be a local penalty function. A penalty function $P : \mathbb{R}^{n+1} \to \mathbb{R}^+$ is defined, for any $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$ as:

$$P(\mathbf{x}, y) = \sum_{i=1}^{n} LP(x_i, y), \tag{1}$$

where $y \in \mathbb{R}$ is called the fused value of $\mathbf{x} \in \mathbb{R}^n$.

The main problem with this approach is that the minimizer y^* may not exist or may be not unique [19]

The following step in this step was done by Calvo et al. [10], proposing the following definition:

Definition 7 The function $LP : \mathbb{I}^2 \to [0, \infty]$ is a local penalty function on $\mathbb{I}$ if and only if, for any $x, y \in \mathbb{I}$, it satisfies:

(LP3.4-1) $LP(x, y) = 0$ if $x = y$, and
(LP3.4-2) $LP(x, y) \geq LP(z, y)$, whenever $x \geq z \geq y$ or $x \leq z \leq y$.

By considering local penalty functions such that

$$LP = K \circ f, \quad \text{that is,} \quad LP(x, y) = K(f(x), f(y)), \tag{2}$$

where $f : \mathbb{I} \to [-\infty, +\infty]$ is a continuous strictly monotonic function and $K : [-\infty, +\infty]^2 \to [0, \infty]$ is a local penalty function, which is convex in each component, Calvo et al. [10] avoided the problem of non-existence of minimizers. This kind of functions were called faithful local penalty functions, and lad to the following definition of penalty function.

Definition 8 Let $LP : \mathbb{I}^2 \to [0, \infty]$ be a faithful local penalty function. A function $f_P : \bigcup_{n \in \mathbb{N}} \mathbb{I}^n \to \mathbb{I}$, defined for all $\mathbf{x} \in \mathbb{I}^n$ and $n \in \mathbb{N}$, by

$$f_P(\mathbf{x}) = \frac{l_{\mathbf{x}} + r_{\mathbf{x}}}{2},$$

where

$$l_{\mathbf{x}} = \inf\{u \in \mathbb{I} \mid \forall v \in \mathbb{I} : P(\mathbf{x}, u) \leq P(\mathbf{x}, v)\}$$
$$r_{\mathbf{x}} = \sup\{u \in \mathbb{I} \mid \forall v \in \mathbb{I} : P(\mathbf{x}, u) \leq P(\mathbf{x}, v)\}$$

is called a penalty-based function, or P-function, for short.

Nevertheless, the next step was to try to recover most relevant aggregation functions, and in this sense, the definition provided by Calvo et al. did not allow to recover idempotent (and hence, averaging) aggregation functions.

This study led to Calvo and Beliakov to propose, in [9], the following definition of a penalty function:

Definition 9 The function $P : \mathbb{I}^{n+1} \to [0, \infty]$ is a penalty function if and only if it satisfies:

(P3.6-1) $P(\mathbf{x}, y) \geq 0$, for all $\mathbf{x}$,y;

(P3.6-2) $P(\mathbf{x}, y) = 0$ if $\mathbf{x} = \mathbf{y}$, and[1]

(P3.6-3) For every fixed $\mathbf{x}$, the set of minimizers of $P(\mathbf{x}, y)$ is either a singleton or an interval.

Definition 10 A penalty-based function $f : \mathbb{I}^n \to \mathbb{I}$ is defined, for all $\mathbf{x} \in X^n$, by

$$f(\mathbf{x}) = \arg\min_{y} P(\mathbf{x}, y), \tag{3}$$

if y is the unique minimizer, and $y = \frac{a+b}{2}$ if the set of minimizers is the interval $]a, b[$ (or $[a, b]$).

We use P-function to shorten the expression penalty function.

Remark 1 In Definition 9, observe that:

1. The condition **(P3.6-1)** is redundant, since this is a consequence of the fact that the range of the function P is $[0, \infty]$.
2. A singleton can be seen as a degenerated interval $\{k\} = [k, k]$.

Definition 9 presents some problems identified below.

Furthermore, in [1], Beliakov and James considered the problem of aggregating some special kind of discontinuous intervals, called non-convex intervals, and propose the following definition:

Definition 11 The function $P : [0, 1]^{n+1} \to [0, \infty]$ is a penalty function if and only if it satisfies:

(P3.8-1) $P(\mathbf{x}, y) = 0$ if $x_i = y$, for all i;

(P3.8-2) $P(\mathbf{x}, y) > 0$ if $x_i \neq y$ for some i, and

(P3.8-3) For every fixed $\mathbf{x}$, the set of minimizers of $P(\mathbf{x}, y)$ is either a singleton or an interval.

The concept of penalty-based function, for a penalty function in the sense of Definition 11, is defined analogously to Definition 10. It is worth to mention that Definitions 9 and 11 are not equivalent.

Furthermore, the restriction of the domain to $[0, 1]^{n+1}$ is not very significant, since the results can be extended to any other bounded interval in a straightforward way.

With this new definition, many averaging functions that are P-functions. For example, the arithmetic mean AM can be generated from

$$P(\mathbf{x}, y) = \frac{\sum_{i=1}^{n} (x_i - y)^2}{n},$$

[1] Observe that the vector $\mathbf{y}$ should be defined as $\mathbf{y} = \underbrace{(y, \ldots, y)}_{n \text{ times}}$, although the authors had not mentioned that in [9].

which is a penalty function in the sense of Definition 11.

But, in fact, it not yet possible to recover every aggregation function as a penalty-based function. So a further modification of the definition is required in order to get such a result. This step was done in [6], obtaining the following result:

Theorem 1 *A function $f : [0,1]^n \to [0,1]$ is a P-function, for a penalty function in the sense of Definition 11, if and only if f is idempotent.*

It follows that each averaging aggregation function is a P-function, for a penalty function in the sense of Definition 11. But it is worth mentioning that the proof of Theorem 1 uses the same argument as that in [9], but with an extra term, which produces a discontinuity for the penalty function.

Regarding monotonicity issues, Wilkin and Beliakov [16] considered also the use of weakly monotonic aggregation functions (see also [7, 13]). In this way, non-monotone operators, such as the mode, can be recovered.

Definition 12 For any closed, nonempty interval $\mathbb{I} \subseteq [-\infty, +\infty]$, the function $P : \mathbb{I}^{n+1} \to \mathbb{R}$ is a penalty function if and only if it satisfies:

(P3.9-1) $P(\mathbf{x}, y) \geq c$, for all $\mathbf{x} \in \mathbb{I}^n, y \in \mathbb{I}$, for some constant $c \in \mathbb{R}$;
(P3.9-2) $P(\mathbf{x}, y) = c$ if and only if all $x_i = y$, for all $i = 1 \dots n$, and
(P3.9-3) P is quasi-convex in y for any $\mathbf{x}$.

All the previous discussed definitions, however, share different difficulties which made its use for applications complicate. For this reason, in [ ], a new definition of penalty functions encompassing most of the advantages of the discussed ones and all the desired properties of these functions was proposed.

Let $\mathbb{I} \subseteq \mathbb{R}$ be a closed interval.

Definition 13 The function $P : \mathbb{I}^{n+1} \to \mathbb{R}^+$ is a penalty function if and only if there exists $c \in \mathbb{R}^+$ such that:

(P4.1-1) $P(\mathbf{x}, y) \geq c$, for all $\mathbf{x} \in \mathbb{I}^n, y \in \mathbb{I}$;
(P4.1-2) $P(\mathbf{x}, y) = c$ if and only if $x_i = y$, for all $i = 1 \dots n$, and
(P4.1-3) P is quasi-convex lower semi-continuous in y for each $\mathbf{x} \in \mathbb{I}^n$.

Definition 14 Let $f : \mathbb{I}^n \to \mathbb{I}$ be a function and P be a penalty function in the sense of Definition 13. Then f is said to be a P-function if, for each $\mathbf{x} \in \mathbb{I}^n$, one has that

$$f(\mathbf{x}) = \frac{a+b}{2} \tag{4}$$

where

$$[a, b] = cl(Minz(P(\mathbf{x}, \cdot)))$$

and $Minz(P(\mathbf{x}, \cdot))$ is the set of minimizers of $P(\mathbf{x}, \cdot)$, that is:

$$Minz(P(\mathbf{x}, \cdot)) = \{y \in \mathbb{I} \mid P(\mathbf{x}, y) \leq P(\mathbf{x}, z), \text{ for each } z \in \mathbb{I}\}, \tag{5}$$

and $cl(S)$ is the closure of $S \subseteq \mathbb{I}$.

Note that the requirement of quasi convexity and lower semicontinuity ensure that the set of minimizers of $P(\mathbf{x}, \cdot)$ is either a singleton or an interval. In particular, we have now the following result [6]:

Theorem 2 *A function $f : \mathbb{I}^n \to \mathbb{I}$ is a P-function in the sense of Definition 14 if and only if f is idempotent.*

It is also worth to say that it is always possible to get a continuous penalty function P_f, for an idempotent function f.

So, finally, we have the following result.

Proposition 3 *Let $A : \mathbb{I}^n \to \mathbb{I}$ be an increasing function. A is an averaging aggregation function if and only if A is a P-function in the sense of Definition 14.*

4 Penalty Functions in a Cartesian Product of Lattices

In the same way as in the real case averaging functions may be constructed by means of penalty functions, an analogous construction can be done for general lattices. In particular, in [3, 8] this construction inpires itsels on the notion faithful penalty functions and it is done using as a first step restricted dissimilarity functions [4, 5]

Definition 15 [5] A mapping $d_R : [0,1]^2 \to [0,1]$ is a restricted dissimilarity function if:

1. $d_R(x, y) = d_R(y, x)$ for every $x, y \in [0,1]$;
2. $d_R(x, y) = 1$ if and only if $x = 0$ and $y = 1$ or $x = 1$ and $y = 0$; that is, $\{x, y\} = \{0, 1\}$;
3. $d_R(x, y) = 0$ if and only if $x = y$;
4. For any $x, y, z \in [0,1]$, if $x \leq y \leq z$, then $d_R(x, y) \leq d_R(x, z)$ and $d_R(y, z) \leq d_R(x, z)$.

Note that distances between fuzzy sets can be defined in terms of restricted dissimilarity functions.

Definition 16 [17] A mapping $\mathscr{D} : \mathscr{F}\mathscr{S}(U)^2 \to [0,1]$ is a distance over $\mathscr{F}\mathscr{S}(U)$ if

1. $\mathscr{D}(A, B) = \mathscr{D}(B, A)$ for every $A, B \in \mathscr{F}\mathscr{S}(U)$;
2. $\mathscr{D}(A, B) = 0$ if and only if $A = B$;
3. $\mathscr{D}(A, B) = 1$ if and only if A and B are complementary crisp sets;
4. if $A \leq A' \leq B' \leq B$, then $\mathscr{D}(A, B) \geq \mathscr{D}(A', B')$.

Along this section, and to avoid possible confusion, we will denote by the letter M an aggregation function.

Theorem 3 [5] *Let M be an aggregation function such that it satisfies*

(A1) $M(x_1, \ldots, x_n) = 1$ *if and only if* $x_1 = \cdots = x_n = 1$*;*
(A2) $M(x_1, \ldots, x_n) = 0$ *if and only if* $x_1 = \cdots = x_n = 0$*,*

and let $d_R : [0,1]^2 \to [0,1]$ *be a restricted dissimilarity function. Then*

$$\mathscr{D}(A,B) = \underset{i=1}{\overset{n}{M}}(d_R(A(u_i), B(u_i)))$$

for all $A, B \in \mathscr{F}\mathscr{S}(U)$ *defines a distance in the sense of Liu.*

In this paper, whenever we speak of distances between fuzzy sets, we mean distances in the sense of Liu [17].

From the point of view of application it is enough to consider as lattices the Cartesian product of finite chains $\mathscr{C}$. Moreover, and since all the finite chains of the same length are isomorphic to each other, we can always assume that we are working with chains of the type $\mathscr{C} = 0 \leq 1 \leq 2 \leq \cdots \leq n-1$. Recall that, if $\mathscr{L}_k = \{\mathscr{C}_1 \times \cdots \times \mathscr{C}_k, \leq, \wedge, \vee\}$ and $a, b \in \mathscr{L}_k$ such that $a \leq b$, every maximal chain joining a and b has the same length.

So the distance between $x, y \in \mathscr{L}$ can be defined as the length of the chain $\mathscr{C}$ with minimal element $a = \wedge(x, y)$ and maximal element $b = \vee(x, y)$, minus one. That is,

$$d(x, y) = length(\mathscr{C}) - 1.$$

This definition is equivalent to the following.

$$d(x, y) = \sum_{i=1}^{m} d_i(x_i, y_i) = \sum_{i=1}^{m} |x_i - y_i| \tag{6}$$

where d_i is the distance in the i-th chain. It is easy to see that Eq. (6) is a distance, called natural distance.

We restrict to chains with supremum and infimum. We need to extend the definition of distance to L-fuzzy sets using the notion of restricted dissimilatity function.

Consider the lattice $\mathscr{L}_m = \{\mathscr{C}_1 \times \cdots \times \mathscr{C}_m, \leq, \wedge, \vee\}$. For each chain $\mathscr{C}_i$ we denote by $\vee(\mathscr{C}_i)$ and $\wedge(\mathscr{C}_i)$ its top and bottom elemnts, respectively. We also denote

$$1_{\mathscr{L}_m} = (\vee(\mathscr{C}_1), \ldots, \vee(\mathscr{C}_m)),$$
$$0_{\mathscr{L}_m} = (\wedge(\mathscr{C}_1), \ldots, \wedge(\mathscr{C}_m)).$$

Definition 17 Take $\mathscr{L}_m = \{\mathscr{C}_1 \times \cdots \times \mathscr{C}_m, \leq, \wedge, \vee\}$. A mapping

$$\delta_R : \mathscr{L}_m \times \mathscr{L}_m \to \mathscr{L}_m$$

is a lattice restricted dissimilarity function if

1. $\delta_R(x, y) = \delta_R(y, x)$ for any $x, y \in \mathscr{L}_m$;
2. $\delta_R(x, y) = 1_{\mathscr{L}_m}$ if and only if for any $i = 1, \ldots, m$,

$$x_i = \vee(\mathscr{C}_i) \text{ and } y_i = \wedge(\mathscr{C}_i),$$
$$\text{or}$$
$$x_i = \wedge(\mathscr{C}_i) \text{ and } y_i = \vee(\mathscr{C}_i);$$

3. $\delta_R(x, y) = 0_{\mathscr{L}_m}$ if and only if $x = y$;
4. If $x \leq y \leq z$ then $\delta_R(x, y) \leq \delta_R(x, z)$ and $\delta_R(y, z) \leq \delta_R(x, z)$.

Proposition 4 *Let each $\delta_{R_i} : \mathscr{C}_i \times \mathscr{C}_i \to \mathscr{C}_i$ be a lattice restricted dissimilarity function. Then the mapping defined as*

$$\delta_R(x, y) = (\delta_{R_1}(x_1, y_1), \dots, \delta_{R_m}(x_m, y_m)) \tag{7}$$

for every $x, y \in \mathscr{L}_m$ is a lattice restricted dissimilarity function.

Let $\mathscr{F}\mathscr{S}(U)^m$ denote the class $\mathbf{A} = (A_1, \dots, A_m)$ with $A_i : U \to \mathscr{C}_i$ such that $\mathbf{A}(u_i) = (A_1(u_i), \dots, A_m(u_i))$ for every $u_i \in U$. Notice that each of the A_i is an L-fuzzy set in the sense of Goguen [11]; i.e., each A_i is a fuzzy set defined over the lattice $\{\mathscr{C}_i, \leq_i, \wedge_i, \vee_i\}$.

Definition 18 Take $\mathscr{L}_m = \{\mathscr{C}_1 \times \cdots \times \mathscr{C}_m, \leq, \wedge, \vee\}$. A mapping

$$\Omega : \mathscr{F}\mathscr{S}(U)^m \times \mathscr{F}\mathscr{S}(U)^m \to \mathscr{L}_m$$

is a lattice distance in $\mathscr{F}\mathscr{S}(U)^m$ if

1. $\Omega(\mathbf{A}, \mathbf{B}) = \Omega(\mathbf{B}, \mathbf{A})$ for every $\mathbf{A}, \mathbf{B} \in \mathscr{F}\mathscr{S}(U)^m$;
2. $\Omega(\mathbf{A}, \mathbf{B}) = 0_{\mathscr{L}_m}$ if and only if $A_i = B_i$ for every $i = 1, \dots, m$;
3. $\Omega(\mathbf{A}, \mathbf{B}) = 1_{\mathscr{L}_m}$ if and only if for every $i = 1, \dots, m$, A_i and B_i are sets such that for every u_j

$$A_i(u_j) = \vee(\mathscr{C}_i) \text{ and } B_i(u_j) = \wedge(\mathscr{C}_i)$$

   or

$$A_i(u_j) = \wedge(\mathscr{C}_i) \text{ and } B_i(u_j) = \vee(\mathscr{C}_i);$$

4. If $\mathbf{A} \leq \mathbf{A}' \leq \mathbf{B}' \leq \mathbf{B}$, then $\Omega(\mathbf{A}, \mathbf{B}) \geq \Omega(\mathbf{A}', \mathbf{B}')$ where $\mathbf{A} = (A_1, \dots, A_m) \leq (A'_1, \dots, A'_m) = \mathbf{A}'$ if $A_i \leq A'_i$ for every i.

we have the nest straight result.

Proposition 5 *Let $M_1, \dots, M_m$ be aggregation functions*

$$M_i : \mathscr{C}_i \times \mathscr{C}_i \to \mathscr{C}_i$$

Then the mapping

$$F : \mathscr{L}_m \times \mathscr{L}_m \to \mathscr{L}_m \text{ given by}$$
$$F(\boldsymbol{x}, \boldsymbol{y}) = (M_1(x_1, y_1), \ldots, M_m(x_m, y_m))$$

is an aggregation function over $\mathscr{L}_m$.

Now we can introduce a method to build lattice distances.

Proposition 6 *Let* $\delta_{R_1}, \ldots, \delta_{R_m}$ *be lattice restricted dissimilarity functions* $\delta_{R_i} : \mathscr{C}_i \times \mathscr{C}_i \to \mathscr{C}_i$. *Let* $M_1, \ldots, M_m$ *be aggregation functions* $M_i : \mathscr{C}_i \times \cdots \times \mathscr{C}_i \to \mathscr{C}_i$ *such that*

(L1) $M_i(x_1, \ldots, x_n) = 1_{\mathscr{L}}$ *if and only if* $x_i = \vee(\mathscr{C}_i)$ *for every* $i = 1, \ldots, n$
(L2) $M_i(x_1, \ldots, x_n) = 0_{\mathscr{L}}$ *if and only if* $x_i = \wedge(\mathscr{C}_i)$ *for every* $i = 1, \ldots, n$

Then

$$\Omega(\boldsymbol{A}, \boldsymbol{B}) = \left(\underset{i=1}{\overset{n}{M_1}} (\delta_{R_1}(A_1(u_i), B_1(u_i))), \ldots, \underset{i=1}{\overset{n}{M_m}} (\delta_{R_m}(A_m(u_i), B_m(u_i))) \right) \tag{8}$$

defines a lattice distance in $\mathscr{F}\mathscr{S}(U)^m$.

We know that the arithmetic mean of convex functions is also a convex function. Next, we consider aggregation functions such that applied to convex functions we obtain another convex function, as in the arithmetic mean case.

Theorem 4 *Let* $Y = (y_1, \ldots, y_m) \in \mathscr{L}_m$. *For each* y_i $(i = 1, \ldots, m)$ *we consider the set*

$$B_{y_i}(u_j) = y_i \text{ for all } u_j \in U \tag{9}$$

and let $\boldsymbol{B}_Y = (B_{y_1}, \ldots, B_{y_m}) \in \mathscr{F}\mathscr{S}(U)^m$. *Let* $M_1, \ldots, M_m$ *be aggregation functions* $M_i : \mathscr{C}_i \times \cdots \times \mathscr{C}_i \to \mathscr{C}_i$ *such that each of them when composed with convex functions is also convex. Take the lattice restricted dissimilarity function* $\delta_R(x, y) = (\delta_{R_1}(x_1, y_1), \ldots, \delta_{R_m}(x_m, y_m))$ *such that each* δ_{R_i} *with* $i = 1, \ldots, m$ *is convex in one variable. Then*

$$P_\Omega : \mathscr{F}\mathscr{S}(U)^{m+1} \to \mathscr{L}_m \text{ given by}$$
$$P_\Omega(\boldsymbol{A}, Y) = \Omega(\boldsymbol{A}, \boldsymbol{B}_Y) = \left(\underset{i=1}{\overset{n}{M_1}} (\delta_{R_1}(A_1(u_i), y_1)), \ldots, \underset{i=1}{\overset{n}{M_m}} (\delta_{R_m}(A_m(u_i), y_m)) \right) \tag{10}$$

satisfies:

1. $P_\Omega(\boldsymbol{A}, Y) \geq 0_{\mathscr{L}_m}$;
2. $P_\Omega(\boldsymbol{A}, Y) = 0_{\mathscr{L}_m}$ *if* $A_k(u_j) = y_k$ *for every* k *and for every* j;
3. *Each of its components is convex with respect to the corresponding* y_k $(k = 1, \ldots, m)$.

Analogously to the case of (see [10, 14]), we introduce the definition of **lattice faithful restricted dissimilarity functions**:

$$\delta_R(x, y) = K(d(x, y)) = K(\sum_{i=1}^{m} |x_i - y_i|) \tag{11}$$

with $K : \mathscr{C} \to \mathscr{C}$ a convex with a unique minimum at $K(0) = 0$.

Theorem 5 *In the setting of Theorem 4, if $\delta_{R_1}, \ldots, \delta_{R_m}$ are lattice faithful restricted dissimilarity functions, then the mapping*

$$F_{\mathscr{L}_m} : \mathscr{F}\mathscr{S}(U)^m \to \mathscr{L}_m \text{ given by}$$
$$F_{\mathscr{L}_m}(\boldsymbol{A}) = \arg\min_Y P_\Omega(\boldsymbol{A}, Y) = \arg\min_Y \Omega(\boldsymbol{A}, \boldsymbol{B}_Y)$$
$$= \left(\arg\min_{y_1} (\mathop{M_1}_{i=1}^{n} (K_1(d(A_1(u_i), y_1))), \ldots, \arg\min_{y_m} (\mathop{M_m}_{i=1}^{n} (K_m(d(A_m(u_i), y_m))) \right)$$
$$= \left(\arg\min_{y_1} (\mathop{M}_{i=1}^{n} (K_1(|A_1(u_i) - y_1|))), \ldots, \arg\min_{y_m} (\mathop{M}_{i=1}^{n} (K_m(|A_m(u_i) - y_m|))) \right)$$

is such that each of its components is an averaging aggregation function over $\mathscr{F}\mathscr{S}(U)$ and $F_{\mathscr{L}_m}(\boldsymbol{A})$ is an averaging aggregation function over the Cartesian product $\mathscr{F}\mathscr{S}(U)^m$.

From now on we will denote by B_{y_q} the fuzzy set over U such that all its membership values are equal to $y_q \in [0, 1]$; that is, $B_{y_q}(u_i) = y_q \in [0, 1]$ for all $u_i \in U$.

Let $Y = (y_1, \ldots, y_m)$ and $\mathbf{B}_Y = (B_{y_1}, \ldots, B_{y_m}) \in \mathscr{F}\mathscr{S}(U)^m$. We will denote by $\mathscr{C}^*$ a chain whose elements belong to $[0, 1]$ and by $\mathscr{L}_m^*$ the product such that $\mathscr{L}_m^* = \mathscr{C}^* \times \cdots \times \mathscr{C}^*$.

Finally, the penalty functions can be obtained as follows.

Theorem 6 *Let $K_i : \mathbb{R} \to \mathbb{R}^+$ be convex functions with a unique minimum at $K_i(0) =$ 0 ($i = 1, \ldots, m$), and take the distance between fuzzy sets defined as*

$$\mathscr{D}(A, B) = \sum_{i=1}^{n} |A(u_i) - B(u_i)| \tag{12}$$

where $A, B \in \mathscr{F}\mathscr{S}(U)$ and $Cardinal(U) = n$. Then the mapping

$$P_\nabla : \mathscr{F}\mathscr{S}(U)^m \times \mathscr{L}_m^* \to \mathbb{R}^+ \text{ given by}$$
$$P_\nabla(\boldsymbol{A}, Y) = \sum_{q=1}^{m} K_q(\mathscr{D}(A_q, B_{y_q})) = \sum_{q=1}^{m} K_q\left(\sum_{p=1}^{n} |A_q(u_p) - y_q|\right) \tag{13}$$

satisfies

1. $P_\nabla(\boldsymbol{A}, Y) \geq 0$;
2. $P_\nabla(\boldsymbol{A}, Y) = 0$ *if and only if* $A_q = y_q$ *for every* $q = 1, \dots, m$;
3. *is convex in* y_q *for every* $q = 1, \dots, m$.

Observe that P_∇ is a penalty function defined over the Cartesian product of lattices $\mathscr{L}_m^{*n+1}$.

Theorem 7 *In the setting of Theorem 6, the mapping*

$$F(\boldsymbol{A}) = \mu = \arg\min_Y P_\nabla(\boldsymbol{A}, Y) \tag{14}$$

where μ *is the rounding to the smallest closest element, is an averaging aggregation function.*

Penalty functions on Cartesian product of lattices have shown themselves very useful in decision making and consensus, see [3, 8].

5 Conclusions

In this chapter we have made a revision of the ideas of penalty and penalty-based functions. We have also discussed how such notions can be extended to deal with data in Cartesian products of lattices.

Acknowledgements This work was supported by research project TIN2016-77356-P(AEI/FEDER, UE)

References

1. Beliakov, G., James, S.: A penalty-based aggregation operator for non-convex intervals. Knowl.-Based Syst. **70**, 335–344 (2014)
2. Beliakov, G., Pradera, A., Calvo, T.: Aggregation Functions: A Guide for Practitioners. Springer, Berlin (2007)
3. Bustince, H., Barrenechea, E., Calvo, T., James, S., Beliakov, G.: Consensus in multi-expert decision making problems using penalty functions defined over a cartesian product of lattices. Inf. Fusion **17**, 56–64 (2014)
4. Bustince, H., Barrenechea, E., Pagola, M.: Image thresholding using restricted equivalence functions and maximizing the measures of similarity. Fuzzy Sets Syst. **158**(5), 496–516 (2007)
5. Bustince, H., Barrenechea, E., Pagola, M.: Relationship between restricted dissimilarity functions, restricted equivalence functions and normal en-functions: Image thresholding invariant. Pattern Recogn. Lett. **29**(4), 525–536 (2008)
6. Bustince, H., Beliakov, G., Dimuro, G.P., Bedregal, B., Mesiar, R.: On the definition of penalty functions in data aggregation. Fuzzy Sets Syst. (2016)

7. Bustince, H., Fernandez, J., Kolesárová, A., Mesiar, R.: Directional monotonicity of fusion functions. Eur. J. Oper. Res. **244**(1), 300–308 (2015)
8. Bustince, H., Jurio, A., Pradera, A., Mesiar, R., Beliakov, G.: Generalization of the weighted voting method using penalty functions constructed via faithful restricted dissimilarity functions. Eur. J. Oper. Res. **225**(3), 472–478 (2013)
9. Calvo, T., Beliakov, G.: Aggregation functions based on penalties. Fuzzy Sets Syst. **161**(10), 1420–1436 (2010)
10. Calvo, T., Mesiar, R., Yager, R.R.: Quantitative weights and aggregation. IEEE Trans. Fuzzy Syst. **12**(1), 62–69 (2004)
11. Goguen, J.: L-fuzzy sets. J. Math. Anal. Appl. **18**(4), 145–174 (1967)
12. Grabisch, M., Marichal, J., Mesiar, R., Pap, E.: Aggregation Functions. Cambridge University Press, Cambridge (2009)
13. Lucca, G., Sanz, J., Dimuro, G., Bedregal, B., Mesiar, R., Kolesárová, A., Bustince, H.: Pre-aggregation functions: construction and an application. IEEE Trans. Fuzzy Syst. (2015)
14. Mesiar, R.: Fuzzy set approach to the utility, preference relations, and aggregation operators. Eur. J. Oper. Res. **176**, 414–422 (2007)
15. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
16. Wilkin, T., Beliakov, G.: Weakly monotonic averaging functions. Int. J. Intell. Syst. **30**(2), 144–169 (2015)
17. Xuecheng, L.: Entropy, distance measure and similarity measure of fuzzy sets and their relations. Fuzzy Sets Syst. **52**, 305–318 (1992)
18. Yager, R.R.: Toward a general theory of information aggregation. Inf. Sci. **68**(3), 191–206 (1993)
19. Yager, R.R., Rybalov, A.: Understanding the median as a fusion operator. Int. J. Gen. Syst. **26**(3), 239–263 (1997)

The Single Parameter Family of Gini Bonferroni Welfare Functions and the Binomial Decomposition, Transfer Sensitivity and Positional Transfer Sensitivity

Silvia Bortot, Mario Fedrizzi, Ricardo Alberto Marques Pereira and Anastasia Stamatopoulou

Abstract We consider the binomial decomposition of generalized Gini welfare functions in terms of the binomial welfare functions $C_j, j = 1, \dots, n$ and we examine the weighting structure of the latter, which progressively focus on the poorest part of the population. In relation with the generalized Gini welfare functions, we introduce measures of transfer sensitivity and positional transfer sensitivity and we illustrate the behaviour of the binomial welfare functions $C_j, j = 1, \dots, n$ with respect to these measures. We investigate the binomial decomposition of the Gini Bonferroni welfare functions and we illustrate the dependence of the binomial decomposition coefficients in relation with the single parameter which describes the family. Moreover we examine the family of Gini Bonferroni welfare functions with respect to the transfer sensitivity and positional transfer sensitivity principles.

Keywords Generalized Gini welfare functions · Binomial decomposition · Single parameter family of Gini Bonferroni welfare functions · Principle of transfer sensitivity · Principle of positional transfer sensitivity

1 Introduction

The study of welfare and inequality has been the research interest of many economical and social scientists, and has been understood as an investigation on the departure from the ideal situation of economic equalitarianism, where each individual of the population has an equal share of the total income. In this sense, different welfare and inequality measures, with different characteristics, have been introduced in the literature in order to express the fairness of the income distribution in society.

S. Bortot(✉) · R.A. Marques Pereira · A. Stamatopoulou
Department of Economics and Management, University of Trento,
Via Inama, 5, 38122 Trento, Italy
e-mail: silvia.bortot@unitn.it

M. Fedrizzi
Department of Industrial Engineering, University of Trento,
Via Sommarive, 9, 38123 Trento, Italy

D.A. Pelta and C. Cruz Corona (eds.), *Soft Computing Based Optimization and Decision Models*, Studies in Fuzziness and Soft Computing 360,
DOI 10.1007/978-3-319-64286-4_18

The generalized Gini welfare functions introduced by Weymark [60], and the associated inequality indices in Atkinson-Kolm-Sen's (AKS) framework, see Atkinson [5], Kolm [48, 49], and Sen [55], are related by Blackorby and Donaldson's correspondence formula [13, 15], $A(\boldsymbol{x}) = \bar{x} - G(\boldsymbol{x})$, where $A(\boldsymbol{x})$ denotes a generalized Gini welfare function, $G(\boldsymbol{x})$ is the associated absolute inequality index, and $\bar{x}$ is the plain mean of the income distribution $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{D}^n$ of a population of $n \geq 2$ individuals, with income domain $\mathbb{D} = [0, \infty)$.

The generalized Gini welfare functions [60] have the form $A(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_{(i)}$ where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ and, as required by the principle of inequality aversion, $w_1 \geq w_2 \geq \cdots \geq w_n \geq 0$ with $\sum_{i=1}^{n} w_i = 1$. These welfare functions correspond to a particular class of the ordered weighted averaging (OWA) functions introduced by Yager [63], which in turn correspond [34] to the Choquet integrals associated with symmetric capacities.

In this paper we recall the binomial decomposition of generalized Gini welfare functions due to Calvo and De Baets [22], see also Bortot and Pereira [20]. The binomial decomposition is formulated in terms of the functional basis formed by the binomial welfare functions.

The binomial welfare functions, denoted C_j with $j = 1, \dots, n$, have null weights associated with the $j - 1$ richest individuals in the population and therefore they are progressively focused on the poorest sector of the population.

The paper is organized as follows. In Sect. 2 we review the notions of generalized Gini welfare function and associated generalized Gini inequality index, and we introduce general measures of transfer sensitivity and positional transfer sensitivity.

In Sect. 3 we briefly review the Gini and Bonferroni welfare functions and inequality indices, and we examine them with respect to the principles of transfer sensitivity and positional transfer sensitivity.

In Sect. 4 we consider the binomial decomposition of generalized Gini welfare functions in terms of the binomial welfare functions $C_j, j = 1, \dots, n$. We illustrate the weights of the binomial welfare functions $C_j, j = 1, \dots, n$, which progressively focus on the poorest sector of the population, and we examine their transfer sensitivity and positional transfer sensitivity properties.

Finally, in Sect. 5 we investigate the Gini Bonferroni welfare functions with parameter $\gamma \in [0, 1]$, particularly in the context of the binomial decomposition. Moreover, we illustrate the weighting structure of the Gini Bonferroni welfare functions and we study their measures of transfer sensitivity and positional transfer sensitivity in terms of the parameter $\gamma \in [0, 1]$. Section 6 contains some conclusive remarks.

2 Generalized Gini Welfare Functions and Inequality Indices

In this section we consider populations of $n \geq 2$ individuals and we briefly review the notions of generalized Gini welfare function and generalized Gini inequality index over the income domain $\mathbb{D} = [0, \infty)$. The income distributions in this framework are represented by points $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n$. We introduce general measures of transfer sensitivity and positional transfer sensitivity.

We begin by presenting notation and basic definitions regarding averaging functions on the domain $\mathbb{D}^n$, with $n \geq 2$ throughout the text. Comprehensive reviews of averaging functions can be found in Chisini [27], Fodor and Roubens [35], Calvo et al. [23], Beliakov et al. [10], Grabisch et al. [46], and Beliakov et al. [9].

Notation. Points in $\mathbb{D}^n$ are denoted $\boldsymbol{x} = (x_1, \dots, x_n)$, with $\mathbf{1} = (1, \dots, 1)$, $\mathbf{0} = (0, \dots, 0)$. Accordingly, for every $x \in \mathbb{D}$, we have $x \cdot \mathbf{1} = (x, \dots, x)$. Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n$, by $\boldsymbol{x} \geq \boldsymbol{y}$ we mean $x_i \geq y_i$ for every $i = 1, \dots, n$, and by $\boldsymbol{x} > \boldsymbol{y}$ we mean $\boldsymbol{x} \geq \boldsymbol{y}$ and $\boldsymbol{x} \neq \boldsymbol{y}$. Given $\boldsymbol{x} \in \mathbb{D}^n$, the increasing and decreasing reorderings of the coordinates of $\boldsymbol{x}$ are indicated as $x_{(1)} \leq \cdots \leq x_{(n)}$ and $x_{[1]} \geq \cdots \geq x_{[n]}$, respectively. In particular, $x_{(1)} = \min\{x_1, \dots, x_n\} = x_{[n]}$ and $x_{(n)} = \max\{x_1, \dots, x_n\} = x_{[1]}$. In general, given a permutation σ on $\{1, \dots, n\}$, we denote $\boldsymbol{x}_\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(n)})$. Finally, the arithmetic mean is denoted $\bar{x} = (x_1 + \cdots + x_n)/n$.

Definition 1 Let $A : \mathbb{D}^n \longrightarrow \mathbb{D}$ be a function. We say that

1. A is *monotonic* if $\boldsymbol{x} \geq \boldsymbol{y} \;\Rightarrow\; A(\boldsymbol{x}) \geq A(\boldsymbol{y})$, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n$. Moreover, A is *strictly monotonic* if $\boldsymbol{x} > \boldsymbol{y} \;\Rightarrow\; A(\boldsymbol{x}) > A(\boldsymbol{y})$, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n$.
2. A is *idempotent* if $A(x \cdot \mathbf{1}) = x$, for all $x \in \mathbb{D}$. On the other hand, A is *nilpotent* if $A(x \cdot \mathbf{1}) = 0$, for all $x \in \mathbb{D}$.
3. A is *symmetric* if $A(\boldsymbol{x}_\sigma) = A(\boldsymbol{x})$, for any permutation σ on $\{1, \dots, n\}$ and all $\boldsymbol{x} \in \mathbb{D}^n$.
4. A is *invariant for translations* if $A(\boldsymbol{x} + t \cdot \mathbf{1}) = A(\boldsymbol{x})$, for all $t \in \mathbb{D}$ and $\boldsymbol{x} \in \mathbb{D}^n$. On the other hand, A is *stable for translations* if $A(\boldsymbol{x} + t \cdot \mathbf{1}) = A(\boldsymbol{x}) + t$, for all $t \in \mathbb{D}$ and $\boldsymbol{x} \in \mathbb{D}^n$.
5. A is *invariant for dilations* if $A(t \cdot \boldsymbol{x}) = A(\boldsymbol{x})$, for all $t \in \mathbb{D}$ and $\boldsymbol{x} \in \mathbb{D}^n$. On the other hand, A is *stable for dilations* if $A(t \cdot \boldsymbol{x}) = t\,A(\boldsymbol{x})$, for all $t \in \mathbb{D}$ and $\boldsymbol{x} \in \mathbb{D}^n$.

The terms positive (negative), increasing (decreasing), and monotonic are used in the weak sense. Otherwise these properties are said to be strict.

We introduce the majorization relation on $\mathbb{D}^n$ and we discuss the concept of income transfer following the approach in Marshall and Olkin [51], focusing on the classical results relating majorization, income transfers, see Marshall and Olkin [51, Chap. 4, Proposition A.1].

Definition 2 The *majorization relation* $\preceq$ on $\mathbb{D}^n$ is defined as follows: given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n$ with $\bar{x} = \bar{y}$, we say that

$$\boldsymbol{x} \preceq \boldsymbol{y} \quad \text{if} \quad \sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)} \qquad k = 1, \dots, n \tag{1}$$

where the case $k = n$ is an equality due to $\bar{x} = \bar{y}$. As usual, we write $\boldsymbol{x} \prec \boldsymbol{y}$ if $\boldsymbol{x} \preceq \boldsymbol{y}$ and not $\boldsymbol{y} \preceq \boldsymbol{x}$, and we write $\boldsymbol{x} \sim \boldsymbol{y}$ if $\boldsymbol{x} \preceq \boldsymbol{y}$ and $\boldsymbol{y} \preceq \boldsymbol{x}$. We say that $\boldsymbol{y}$ majorizes $\boldsymbol{x}$ if $\boldsymbol{x} \prec \boldsymbol{y}$, and we say that $\boldsymbol{x}$ and $\boldsymbol{y}$ are indifferent if $\boldsymbol{x} \sim \boldsymbol{y}$.

The majorization relation is a partial preorder, in the sense that $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n$ are comparable only when $\bar{x} = \bar{y}$, and $\boldsymbol{x} \sim \boldsymbol{y}$ if and only if $\boldsymbol{x}$ and $\boldsymbol{y}$ differ by a permutation. Given an income distribution $\boldsymbol{x} \in \mathbb{D}^n$, with mean income $\bar{x}$, it holds that $\bar{x} \cdot \mathbf{1} \preceq \boldsymbol{x}$ since $k\bar{x} \geq \sum_{i=1}^{k} x_{(i)}$ for $k = 1, \ldots, n$. The majorization is strict, $\bar{x} \cdot \mathbf{1} \prec \boldsymbol{x}$, when $\boldsymbol{x}$ is not a uniform income distribution.

Definition 3 Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n$ with $\bar{x} = \bar{y}$, we say that $\boldsymbol{x}$ is derived from $\boldsymbol{y}$ by means of an *income transfer* if, for some pair $i, j = 1, \ldots, n$ with $y_i \leq y_j$, we have

$$x_i = (1 - \varepsilon)\, y_i + \varepsilon y_j \qquad x_j = \varepsilon y_i + (1 - \varepsilon)\, y_j \qquad \varepsilon \in [0, 1] \tag{2}$$

and $x_k = y_k$ for $k \neq i, j$. These formulas express an income transfer, from a richer to a poorer individual, of an income amount $\varepsilon(y_j - y_i)$. The income transfer obtains $\boldsymbol{x} = \boldsymbol{y}$ if $\varepsilon = 0$, and exchanges the relative positions of donor and recipient in the income distribution if $\varepsilon = 1$, in which case $\boldsymbol{x} \sim \boldsymbol{y}$. In the intermediate cases $\varepsilon \in (0, 1)$ the income transfer produces an income distribution $\boldsymbol{x}$ which is majorized by the original $\boldsymbol{y}$, that is $\boldsymbol{x} \prec \boldsymbol{y}$.

In general, for the majorization relation $\preceq$ and income distributions $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n$ with $\bar{x} = \bar{y}$, it holds that $\boldsymbol{x} \preceq \boldsymbol{y}$ if and only if $\boldsymbol{x}$ can be derived from $\boldsymbol{y}$ by means of a finite sequence of income transfers. Moreover, $\boldsymbol{x} \prec \boldsymbol{y}$ if any of the income transfers is not a permutation.

Definition 4 A function $A : \mathbb{D}^n \longrightarrow \mathbb{D}$ is an *averaging function* if it is monotonic and idempotent. An averaging function is said to be *strict* if it is strictly monotonic. Note that monotonicity and idempotency implies that $\min(\boldsymbol{x}) \leq A(\boldsymbol{x}) \leq \max(\boldsymbol{x})$, for all $\boldsymbol{x} \in \mathbb{D}^n$.

Particular instances of averaging functions are weighted averaging (WA) functions, ordered weighted averaging (OWA) functions, and Choquet integrals. The former two are special cases of Choquet integration.

Definition 5 Given a weighting vector $\boldsymbol{w} = (w_1, \ldots, w_n) \in [0, 1]^n$, with $\sum_{i=1}^{n} w_i = 1$, the *Weighted Averaging (WA) function* associated with $\boldsymbol{w}$ is the averaging function $A : \mathbb{D}^n \longrightarrow \mathbb{D}$ defined as

$$A(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_i. \tag{3}$$

Definition 6 Given a weighting vector $\boldsymbol{w} = (w_1, \ldots, w_n) \in [0, 1]^n$, with $\sum_{i=1}^{n} w_i = 1$, the *Ordered Weighted Averaging (OWA) function* associated with $\boldsymbol{w}$ is the averaging function $A : \mathbb{D}^n \longrightarrow \mathbb{D}$ defined as

$$A(\boldsymbol{w}) = \sum_{i=1}^{n} w_i x_{(i)}. \tag{4}$$

The traditional form of OWA functions as introduced by Yager [63] is as follows, $A(\boldsymbol{x}) = \sum_{i=1}^{n} \tilde{w}_i x_{[i]}$ where $\tilde{w}_i = w_{n-i+1}$. In [64, 65] the theory and applications of OWA functions are discussed in detail. The following is a classical result particularly relevant in our framework. This result regards a form of dominance relation between OWA functions and the associated weighting structures, see for instance Bortot and Pereira [20] and references therein.

A class of welfare functions which plays a central role in this paper is that of the generalized Gini welfare functions introduced by Weymark [60], see also Mehran [52], Donaldson and Weymark [30, 31], Yaari [61, 62], Ebert [33], Quiggin [54], Ben-Porath and Gilboa [11].

Definition 7 Given a weighting vector $\boldsymbol{w} = (w_1, \ldots, w_n) \in [0,1]^n$, with $w_1 \geq \cdots \geq w_n \geq 0$ and $\sum_{i=1}^{n} w_i = 1$, the *generalized Gini welfare function* associated with $\boldsymbol{w}$ is the function $A : \mathbb{D}^n \longrightarrow \mathbb{D}$ defined as

$$A(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_{(i)} \tag{5}$$

and, in the AKS framework, the associated *generalized Gini inequality index* is defined as

$$G(\boldsymbol{x}) = \bar{x} - A(\boldsymbol{x}) = -\sum_{i=1}^{n} (w_i - \frac{1}{n}) x_{(i)} . \tag{6}$$

Generalized Gini welfare functions are *strict* if and only if $w_1 > \cdots > w_n > 0$. Moreover, generalized Gini welfare functions are stable for translations and the associated generalized Gini inequality indices are invariant for translations. Both are stable for dilations.

In relation with generalized Gini welfare functions, the principles of transfer sensitivity (TS) and positional transfer sensitivity (PTS) are based on the central notion of a progressive income transfer. Given an income distribution

$$\boldsymbol{x} = \left(x_{(1)}, \ldots, x_{(i)}, \ldots, x_{(j)}, \ldots, x_{(n)}\right)$$

and $i < j$ and $x_{(i)} \leq x_{(j)}$, we consider the progressive transfer of an income amount δ from $x_{(j)}$ to $x_{(i)}$, such that $x_{(i)} + \delta \leq x_{(j)} - \delta$. This progressive transfer results in a new income distribution

$$\boldsymbol{x}' = \left(x_{(1)}, \ldots, x_{(i)} + \delta, \ldots, x_{(j)} - \delta, \ldots, x_{(n)}\right) .$$

We consider thus a progressive income transfer δ from $x_{(j)}$ to $x_{(i)}$ with $i < j$. This transfer results in a new income distribution in which $x'_{(i)} = x_{(i)} + \delta$, $x'_{(j)} = x_{(j)} - \delta$, and $x'_{(k)} = x_{(k)}$ for $k \neq i, j$. From the definition (5) of generalized Gini welfare functions, we obtain

$$
\begin{aligned}
A(\boldsymbol{x}') - A(\boldsymbol{x}) &= \sum_{k=1}^{n} w_k x'_{(k)} - \sum_{k=1}^{n} w_k x_{(k)} \\
&= \Big[\big(w_1 x_{(1)} + \cdots + w_i (x_{(i)} + \delta) + \cdots + w_j (x_{(j)} - \delta) + \cdots + w_n x_{(n)} \big) \\
&\quad - \big(w_1 x_{(1)} + \cdots + w_i x_{(i)} + \cdots + w_j x_{(j)} + \cdots + w_n x_{(n)} \big) \Big] \\
&= \big(w_i - w_j \big) \delta \,. \qquad (7)
\end{aligned}
$$

Given that the weight difference $w_i - w_j$ is non negative, the generalized Gini welfare of the distribution $\boldsymbol{x}'$ is greater or equal than that of the original distribution $\boldsymbol{x}$. This means that the generalized Gini welfare function A satisfies the transfer sensitivity (TS) principle, or Pigou-Dalton principle, which states that welfare (inequality) measures should be non decreasing (non increasing) under progressive income transfers.

On the other hand, the principle of positional transfer sensitivity (PTS) states that the effect of an income transfer generates higher welfare values when it occurs at lower income levels. In fact, the non negative weight difference $w_i - w_j$ can vary with the position indicated by the indices i, j. In particular, with $j = i + 1$, we may have constant weight differences (the classical Gini case) or decreasing weight differences (the classical Bonferroni case), as we will see below.

We can measure the transfer sensitivity of generalized Gini welfare functions $A(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_{(i)}$ by means of

$$
TS(A) = \sum_{i=1}^{n-1} w_i - w_{i+1} = w_1 - w_n \quad \in [0, 1] \qquad (8)
$$

where w_i are the weights of the generalized Gini welfare function, with $i = 1, \ldots, n$.

The TS measure takes values in the unit interval $[0, 1]$. A TS value further away from zero indicates a higher level of transfer sensitivity. More specifically, as the value of the TS measure increases, transfer sensitivity increases too.

We can measure the positional transfer sensitivity of generalized Gini welfare functions $A(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_{(i)} \neq \bar{x}$ by means of

$$
PTS(A) = 1 - \sum_{i=1}^{n-1} \frac{\omega_i \ln \omega_i}{\ln(1/(n-1))} \quad \in [0, 1] \qquad (9)
$$

where ω_i with $i = 1, \ldots, n-1$ is given by

$$
\omega_i = \frac{w_i - w_{i+1}}{w_1 - w_n} \qquad i = 1, \ldots, n-1 \qquad (10)
$$

with $\omega_1, \ldots, \omega_{n-1} \geq 0$ and $\omega_1 + \cdots + \omega_{n-1} = 1$. In the case $\omega = 0$ we conventionally take $\omega \ln \omega = 0$.

This measure takes values in the unit interval $[0, 1]$. In fact, the summation term in (9), corresponding to the Shannon entropy of the $\omega_1, \ldots, \omega_{n-1}$ distribution, takes values in [0, 1] and reaches the maximum value 1 when such distribution is uniform, $\omega_1 = \cdots = \omega_{n-1} = 1/(n-1)$. Therefore, the higher the value of the $PTS(A)$ measure, the greater the positional transfer sensitivity of generalized Gini welfare function A in relation with income transfers from individual $j+1$ to individual j, with $j = 1, \ldots, n$.

3 Gini and Bonferroni Welfare Functions and the Associated inequality Indices

The classical Gini [37–39], Bonferroni [18, 19], and De Vergottini [28, 29] welfare functions and the associated inequality indices are classical instances of the AKS generalized Gini framework. In this section we recall the basic facts about the Gini and Bonferroni welfare functions and inequality indices and we examine their properties regarding transfer sensitivity and positional transfer sensitivity.

The classical Gini welfare function $A_G(\boldsymbol{x})$ and the associated classical Gini inequality index $G(\boldsymbol{x}) = \bar{x} - A_G(\boldsymbol{x})$ are defined as

$$A_G(\boldsymbol{x}) = \sum_{i=1}^{n} w_i^G x_{(i)} \qquad w_i^G = \frac{2(n-i)+1}{n^2} \tag{11}$$

where the weights of $A_G(\boldsymbol{x})$ are positive and strictly decreasing with unit sum, $\sum_{i=1}^{n} w_i^G = 1$, and

$$G(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\frac{1}{n} - w_i^G\right) x_{(i)} = -\sum_{i=1}^{n} \frac{n-2i+1}{n^2} x_{(i)} \tag{12}$$

where the coefficients of $G(\boldsymbol{x})$ have zero sum.

The classical absolute Gini inequality index G is traditionally defined as

$$G(\boldsymbol{x}) = \frac{1}{2n^2} \sum_{i,j=1}^{n} |x_i - x_j| = -\frac{1}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(x_{(i)} - x_{(j)}\right) \tag{13}$$

where the double summation expression for $-n^2 G(\boldsymbol{x})$ in (13) can be written as

$$(-(n-1))x_{(1)} + (1-(n-2))x_{(2)} + \cdots + ((n-2)-1)x_{(n-1)} + (n-1)x_{(n)} \tag{14}$$

which corresponds to (12).

The classical Bonferroni welfare function $A_B(\boldsymbol{x})$ and the associated classical Bonferroni inequality index $B(\boldsymbol{x}) = \bar{x} - A_B(\boldsymbol{x})$ are defined as

$$A_B(\boldsymbol{x}) = \sum_{i=1}^{n} w_i^B x_{(i)} \qquad\qquad w_i^B = \sum_{j=i}^{n} \frac{1}{jn} \tag{15}$$

where the weights of $A_B(\boldsymbol{x})$ are positive and strictly decreasing with unit sum, $\sum_{i=1}^{n} w_i^B = 1$, and

$$B(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\frac{1}{n} - w_i^B\right) x_{(i)} \tag{16}$$

where the coefficients of $B(\boldsymbol{x})$ have zero sum.

The classical absolute Bonferroni inequality index B is traditionally defined as

$$B(\boldsymbol{x}) = \bar{x} - \frac{1}{n} \sum_{i=1}^{n} m_i(\boldsymbol{x}) \tag{17}$$

where the mean income of the i poorest individuals in the population is given by

$$m_i(\boldsymbol{x}) = \frac{1}{i} \sum_{j=1}^{i} x_{(j)} \qquad \text{for} \quad i = 1, \ldots, n\,. \tag{18}$$

Therefore we have

$$\begin{aligned}
A_B(\boldsymbol{x}) &= \frac{1}{n} \sum_{i=1}^{n} m_i(\boldsymbol{x}) &(19)\\
&= \frac{1}{n}\left[\left(x_{(1)}\right) + \frac{1}{2}\left(x_{(1)} + x_{(2)}\right) + \cdots + \frac{1}{n}\left(x_{(1)} + \cdots + x_{(n)}\right)\right] &(20)\\
&= \frac{1}{n}\left[\sum_{j=1}^{n} \frac{1}{j} x_{(1)} + \sum_{j=2}^{n} \frac{1}{j} x_{(2)} + \cdots \sum_{j=n}^{n} \frac{1}{j} x_{(n)}\right] &(21)
\end{aligned}$$

which corresponds to (15).

The rich literature on the three classical cases of generalized Gini welfare functions—Gini, Bonferroni and De Vergottini—includes, for instance, Kolm [47], Atkinson [5], Sen [55, 56], Mehran [52], Blackorby and Donaldson [13–16], Lorenzen [50], Donaldson and Weymark [30, 31], Nygård and Sandström [53], Blackorby et al. [17], Weymark [60], Yitzhaki [66], Giorgi [40, 41], Benedetti [12], Ebert [32], Shorrocks and Foster [57], Yaari [62], Silber [58], Bossert [21], Tarsitano [59], Ben Porath and Gilboa [11], Zoli [68], Gajdos [36], Aaberge [1–3], Giorgi and Crescenzi [42, 43], Chakravarty and Muliere [26], Chakravarty [24, 25], Bárcena and Imedio [6], Giorgi and Nadarajah [44], Bárcena and Silber [7, 8], Aristondo et al. [4], and Zenga [67].

We now consider a progressive transfer δ from $x_{(j)}$ to $x_{(i)}$ with $i < j$. This transfer results in a new income distribution in which $x'_{(i)} = x_{(i)} + \delta$, $x'_{(j)} = x_{(j)} - \delta$, and $x'_{(k)} = x_{(k)}$ for $k \neq i, j$. From (11) and (15) we obtain

$$
\begin{aligned}
A_G(\boldsymbol{x}') - A_G(\boldsymbol{x}) &= \sum_{k=1}^{n} w_k^G x'_{(k)} - \sum_{k=1}^{n} w_k^G x_{(k)} \\
&= \Big[\big(w_1^G x_{(1)} + \cdots + w_i^G (x_{(i)} + \delta) + \cdots + w_j^G (x_{(j)} - \delta) + \cdots + w_n^G x_{(n)} \big) \\
&\quad - \big(w_1^G x_{(1)} + \cdots + w_i^G x_{(i)} + \cdots + w_j^G x_{(j)} + \cdots + w_n^G x_{(n)} \big) \Big] \\
&= \big(w_i^G \delta - w_j^G \delta \big) = \left(\frac{2(n-i)+1}{n^2} - \frac{2(n-j)+1}{n^2} \right) \delta \\
&= \frac{2}{n^2} (j-i)\, \delta
\end{aligned}
\tag{22}
$$

$$
\begin{aligned}
A_B(\boldsymbol{x}') - A_B(\boldsymbol{x}) &= \sum_{k=1}^{n} w_k^B x'_{(k)} - \sum_{k=1}^{n} w_k^B x_{(k)} \\
&= \Big[\big(w_1^B x_{(1)} + \cdots + w_i^B (x_{(i)} + \delta) + \cdots + w_j^B (x_{(j)} - \delta) + \cdots + w_n^B x_{(n)} \big) \\
&\quad - \big(w_1^B x_{(1)} + \cdots + w_i^B x_{(i)} + \cdots + w_j^B x_{(j)} + \cdots + w_n^B x_{(n)} \big) \Big] \\
&= \big(w_i^B \delta - w_j^B \delta \big) = \left(\sum_{k=i}^{n} \frac{1}{nk} - \sum_{k=j}^{n} \frac{1}{nk} \right) \delta \\
&= \frac{1}{n} \left(\frac{1}{i} + \frac{1}{i+1} + \cdots + \frac{1}{j-1} \right) \delta = \left(\frac{1}{n} \sum_{k=i}^{j-1} \frac{1}{k} \right) \delta.
\end{aligned}
\tag{23}
$$

Since $A_G(\boldsymbol{x}') - A_G(\boldsymbol{x}) > 0$ and $A_B(\boldsymbol{x}') - A_B(\boldsymbol{x}) > 0$, both welfare functions satisfy the principle of transfer sensitivity. Expression (22) implies that the increase in welfare, in the Gini case, depends on the difference $(j-i)$, irrespectively of the actual positions i, j. The Bonferroni welfare function, on the other hand, does depend on the actual positions i, j. Expression (23) indicates that the increase in welfare is greater if the transfer occurs at lower income levels and therefore the Bonferroni welfare function satisfies the principle of positional transfer sensitivity.

4 The Binomial Decomposition

In this section we review the binomial decomposition of generalized Gini welfare functions due to Calvo and De Baets [22] and Bortot and Pereira [20]. We examine the weighting structures of the binomial welfare functions C_j, with $j = 1, \ldots, n$, and we illustrate their properties regarding transfer sensitivity and positional transfer sensitivity.

Definition 8 The *binomial welfare functions* $C_j : \mathbb{D}^n \longrightarrow \mathbb{D}$, with $j = 1, \ldots, n$, are defined as

$$C_j(\boldsymbol{x}) = \sum_{i=1}^{n} w_{ji}\, x_{(i)} = \sum_{i=1}^{n} \frac{\binom{n-i}{j-1}}{\binom{n}{j}}\, x_{(i)} \qquad j = 1, \ldots, n \tag{24}$$

where the binomial weights w_{ji}, $i, j = 1, \ldots, n$ are null when $i + j > n + 1$, according to the usual convention that $\binom{p}{q} = 0$ when $p < q$, with $p, q = 0, 1, \ldots$ Given that the binomial weights are decreasing, $w_{j1} \geq w_{j2} \geq \cdots \geq w_{jn}$ for $j = 1, \ldots, n$, the binomial welfare functions are generalized Gini welfare functions.

With the exception of $C_1(\boldsymbol{x}) = \bar{x}$, the binomial welfare functions C_j, $j = 2, \ldots, n$ have an increasing number of null weights, in correspondence with $x_{(n-j+2)}, \ldots, x_{(n)}$. The weight normalization of the binomial welfare functions, $\sum_{i=1}^{n} w_{ji} = 1$ for $j = 1, \ldots, n$, is due to the column-sum property of binomial coefficients,

$$\sum_{i=1}^{n} \binom{n-i}{j-1} = \sum_{i=0}^{n-1} \binom{i}{j-1} = \binom{n}{j} \qquad j = 1, \ldots, n\,. \tag{25}$$

The binomial welfare functions C_j, $j = 1, \ldots, n$ are continuous, idempotent, and stable for translations, where the latter two properties follow immediately from the unit sum normalization of the binomial weights. Moreover, due to the cumulative property of the binomial weights, see Calvo and De Baets [22], see also Bortot and Pereira [20], the binomial welfare functions satisfy the relations $\bar{x} = C_1(\boldsymbol{x}) \geq C_2(\boldsymbol{x}) \geq \cdots \geq C_n(\boldsymbol{x}) \geq 0$, for any $\boldsymbol{x} \in \mathbb{D}^n$.

Proposition 1 *Generalized Gini welfare functions* $A : \mathbb{D}^n \longrightarrow \mathbb{D}$ *can be written uniquely as*

$$A(\boldsymbol{x}) = \alpha_1 C_1(\boldsymbol{x}) + \alpha_2 C_2(\boldsymbol{x}) + \cdots + \alpha_n C_n(\boldsymbol{x}) \tag{26}$$

where the coefficients α_j, $j = 1, \ldots, n$ *are subject to the following conditions,*

$$\alpha_1 = 1 - \sum_{j=2}^{n} \alpha_j \geq 0 \tag{27}$$

$$\sum_{j=2}^{n} \left[1 - n \frac{\binom{i-1}{j-1}}{\binom{n}{j}} \right] \alpha_j \leq 1 \qquad i = 2, \ldots, n \tag{28}$$

$$\sum_{j=2}^{n} \frac{\binom{n-i}{j-2}}{\binom{n}{j}}\, \alpha_j \geq 0 \qquad i = 2, \ldots, n\,. \tag{29}$$

The binomial welfare functions constitute therefore a functional basis for the generalized Gini welfare functions, which can be uniquely expressed as

$A(\mathbf{x}) = \sum_{j=1}^{n} \alpha_j\, C_j(\mathbf{x})$ where the coefficients α_j, $j = 1, \dots, n$ satisfy the constraints (27)–(29), one of which is $\sum_{j=1}^{n} \alpha_j = 1$. However, the binomial decomposition does not express a simple convex combination of the binomial welfare functions, as the condition $\alpha_1 + \cdots + \alpha_n = 1$ might suggest. In fact, condition (27) ensures $\alpha_1 \geq 0$ but conditions (28) and (29) allow for negative $\alpha_2, \dots, \alpha_n$ values.

Notice that $C_1(\boldsymbol{x}) = \bar{x}$ and $C_2(\boldsymbol{x})$, which has $n-1$ positive linearly decreasing weights and one null last weight, is the only strict binomial welfare function. In terms of the classical Gini welfare function we have that

$$A^c(\boldsymbol{x}) = \frac{1}{n}\, C_1(\boldsymbol{x}) + \frac{n-1}{n}\, C_2(\boldsymbol{x})\,. \tag{30}$$

The remaining binomial welfare functions $C_j(\boldsymbol{x})$, $j = 3, \dots, n$, have $n - j + 1$ positive non-linear decreasing weights and $j - 1$ null last weights.

In dimensions $n = 4, 8$ the weights $w_{ij} \in [0, 1]$, $i, j = 1, \dots, n$ of the binomial welfare functions C_j, $j = 1, \dots, n$ are as follows,

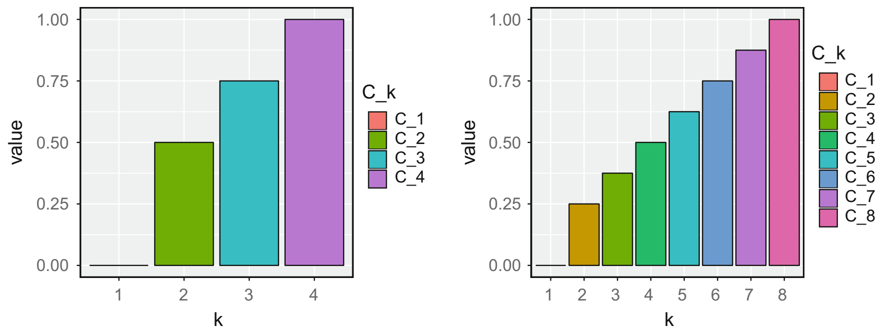
$$\begin{array}{lll} n = 4 & C_1 : (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) \\ & C_2 : (\frac{3}{6}, \frac{2}{6}, \frac{1}{6}, 0) \\ & C_3 : (\frac{3}{4}, \frac{1}{4}, 0, 0) \\ & C_4 : (1, 0, 0, 0) \end{array} \qquad \begin{array}{ll} n = 8 & C_1 : (\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}) \\ & C_2 : (\frac{7}{28}, \frac{6}{28}, \frac{5}{28}, \frac{4}{28}, \frac{3}{28}, \frac{2}{28}, \frac{1}{28}, 0) \\ & C_3 : (\frac{21}{56}, \frac{15}{56}, \frac{10}{56}, \frac{6}{56}, \frac{3}{56}, \frac{1}{56}, 0, 0) \\ & C_4 : (\frac{35}{70}, \frac{20}{70}, \frac{10}{70}, \frac{4}{70}, \frac{1}{70}, 0, 0, 0) \\ & C_5 : (\frac{35}{56}, \frac{15}{56}, \frac{5}{56}, \frac{1}{56}, 0, 0, 0, 0) \\ & C_6 : (\frac{21}{28}, \frac{6}{28}, \frac{1}{28}, 0, 0, 0, 0, 0) \\ & C_7 : (\frac{7}{8}, \frac{1}{8}, 0, 0, 0, 0, 0, 0) \\ & C_8 : (1, 0, 0, 0, 0, 0, 0, 0) \end{array}$$

The binomial welfare functions C_j, $j = 1, \dots, n$ have null weights associated with the $j - 1$ richest individuals in the population and therefore, as j increases from 1 to n, they behave in analogy with poverty measures which progressively focus on the poorest part of the population.

In order to measure the transfer sensitivity of the binomial welfare functions C_j, with $j = 1, \dots, n$ we consider a transfer from the richest to the poorest individual. To measure the transfer sensitivity of the binomial welfare functions C_j, with $j = 1, \dots, n$, we use expression (8).

In Fig. 1 we can see the values of the $TS(C_j)$ measure of the binomial welfare functions C_j, with $j = 1, \dots, n$ for the cases $n = 4, 8$.

In both cases $n = 4, 8$ we observe that TS increases linearly for $j = 3, \dots, n$, which means that the TS difference $C_j - C_{j-1}$ between any 2 consecutive binomial welfare functions is the same. This can be proved as follows,



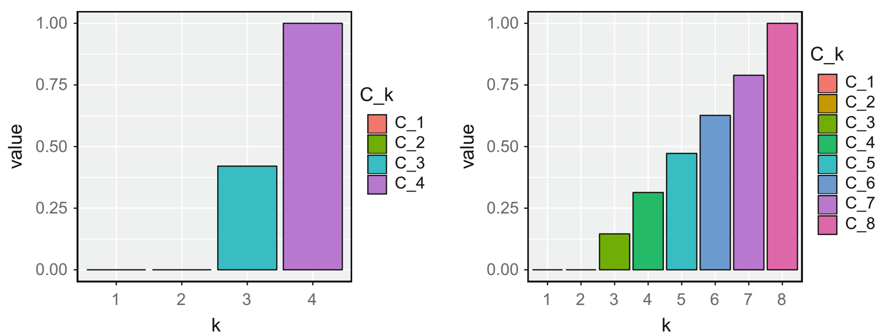
(a) Case $n = 4$ (b) Case $n = 8$

Fig. 1 Transfer sensitivity of C_j, for $j = 1, \dots, n$

$$
\begin{aligned}
(w_{j1} - w_{jn}) - (w_{j-1,1} - w_{j-1,n}) &= \frac{1}{\binom{n}{j}}\left[\binom{n-1}{j-1} - \binom{n-n}{j-1}\right] - \\
&\quad \frac{1}{\binom{n}{j-1}}\left[\binom{n-1}{j-2} - \binom{n-n}{j-2}\right] \\
&= \frac{(n-1)!\,j!}{(j-1)!\,n!} - \frac{(n-1)!(j-1)!}{n!(j-2)!} \\
&= \frac{j}{n} - \frac{j-1}{n} = \frac{1}{n}
\end{aligned}
$$

where w_{ji} are the binomial weights in (24) with $i, j = 1, \dots, n$.

In order to measure the positional transfer sensitivity of the binomial welfare functions C_j, with $j = 1, \dots, n$ we consider $n - 1$ income transfers, each time from an individual in position j to the individual in position $j - 1$, with $j = 1, \dots, n$. To measure the positional transfer sensitivity of the binomial welfare functions C_j, with $j = 1, \dots, n$ we use expression (9)



(a) Case $n = 4$ (b) Case $n = 8$

Fig. 2 Positional transfer sensitivity of C_j, for $j = 1, \dots, n$

In Fig. 2 we illustrate the *PTS* values of the binomial welfare functions C_j, with $j = 1, \ldots, n$ in the cases $n = 4, 8$. We observe in both cases that *PTS* is null for $j = 1, 2$ while for $j = 3, \ldots, n$ it increases monotonically, not linearly.

5 The Single Parameter Gini Bonferroni Welfare Functions

The single parameter family of Gini Bonferroni (GB) welfare functions, which interpolates between the classical Gini and Bonferroni cases, has been introduced by Bárcena and Silber [8]. We recall the definition of the single parameter GB welfare functions and we examine their binomial decomposition. Moreover, we study the measures of transfer sensitivity and positional transfer sensitivity in terms of the parameter $\gamma \in [0, 1]$.

The welfare functions of the GB family are of the form

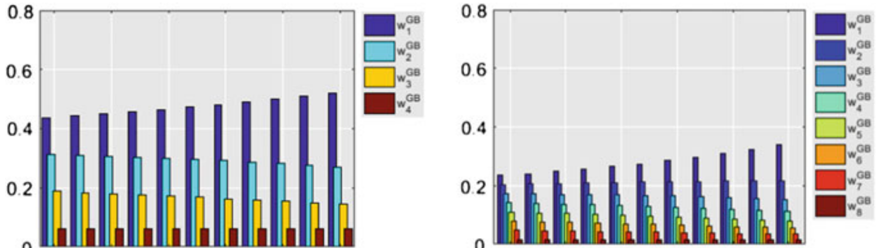
$$A_{GB}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i^{GB} x_{(i)} \tag{31}$$

with

$$w_i^{GB} = (1/n^2)\left[n - i(n/i)^{\gamma} + \sum_{j=i}^{n}(n/j)^{\gamma}\right] \qquad \gamma \in [0, 1] \tag{32}$$

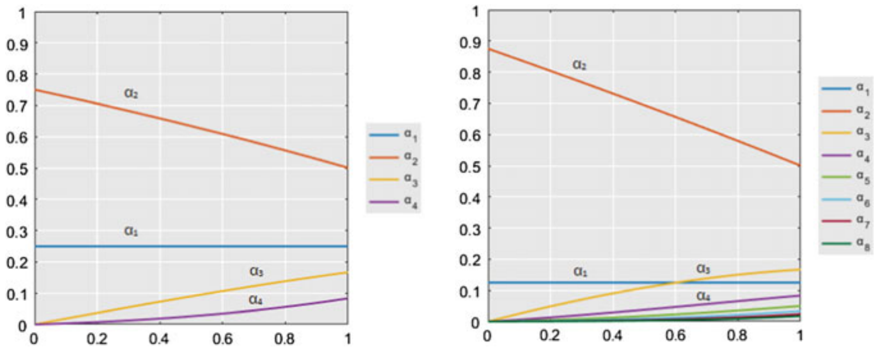
where the classical Gini and Bonferroni welfare functions are special cases with $\gamma = 0, 1$. Note that when $\gamma = 0$ we obtain the "equally distributed equivalent level of income" corresponding to the Gini welfare function, while when $\gamma = 1$ we obtain the "equally distributed equivalent level of income" corresponding to the Bonferroni welfare function.

Given that the weights of the GB welfare functions are strictly decreasing, $w_1^{GB} > w_2^{GB} > \cdots > w_n^{GB} = 1/n^2$, the GB welfare functions are generalized Gini welfare functions. The weighting structure of the GB welfare functions is illustrated in Fig. 3 in the cases $n = 4, 8$.



(a) Case $n = 4$ (b) Case $n = 8$

Fig. 3 Weights of the GB welfare functions for parameter values $\gamma = 0, 0.1, \ldots, 1$



(a) $\alpha_1, ..., \alpha_4$ with $\gamma \in [0,1]$. (b) $\alpha_1, ..., \alpha_8$ with $\gamma \in [0,1]$.

Fig. 4 Coefficients of the binomial decomposition for $n = 4, 8$

In the framework of the binomial decomposition (26), each GB welfare function $A_{GB}(\boldsymbol{x})$ can be uniquely expressed in terms of the binomial Gini welfare functions $C_1, C_2, \dots C_n$ as follows,

$$A_{GB}(\boldsymbol{x}) = \alpha_1 C_1(\boldsymbol{x}) + \alpha_2 C_2(\boldsymbol{x}) + \cdots + \alpha_n C_n(\boldsymbol{x}) \qquad \gamma \in [0,1] \tag{33}$$

which can be written as

$$\sum_{i=1}^{n} w_i^{GB} x_{(i)} = \alpha_1 \sum_{i=1}^{n} w_{1i}\, x_{(i)} + \alpha_2 \sum_{i=1}^{n} w_{2i}\, x_{(i)} + \cdots + \alpha_n \sum_{i=1}^{n} w_{ni}\, x_{(i)} \qquad \gamma \in [0,1]\,. \tag{34}$$

The expression of the binomial decomposition is unique and therefore, for each value of the parameter $\gamma \in [0,1]$, we obtain a unique vector $(\alpha_1, \dots, \alpha_n)$ by solving the linear system

$$\begin{cases} w_1^{GB} = \alpha_1 w_{11} + \alpha_2 w_{21} + \cdots + \alpha_n w_{n1} \\ w_2^{GB} = \alpha_1 w_{12} + \alpha_2 w_{22} + \cdots + \alpha_n w_{n2} \\ \qquad \cdots \\ w_n^{GB} = \alpha_1 w_{1n} + \alpha_2 w_{2n} + \cdots + \alpha_n w_{nn} \end{cases} \tag{35}$$

where the binomial weights w_{ji}, $i,j = 1, \dots, n$ are as in (24).

In Fig. 4 we depict the vector $(\alpha_1, \dots, \alpha_n)$ as a function of the parameter $\gamma \in [0,1]$ in the cases $n = 4, 8$.

We observe, as expected, that $\alpha_1 = 1/n$ is independent of the parameter $\gamma \in [0,1]$ since, in the last equation of the linear system (35), we have $w_n^{GB} = 1/n^2$ and $w_{1n} = 1/n$ and $w_{2n} = \cdots = w_{nn} = 0$.

On the other hand, we observe that only α_2 is decreasing, whereas $\alpha_3, \dots \alpha_n$ are increasing with respect to $\gamma \in [0,1]$.

It is well known that the classical Gini welfare function is 2-additive, see for instance Grabisch [45] and Bortot and Pereira [20] and references therein. On the other hand, the classical Bonferroni welfare function is n-additive. In fact in Fig. 4 we observe that only $\alpha_1, \alpha_2 \neq 0$ in the classical Gini case $\gamma = 0$, and $\alpha_1, \ldots, \alpha_n \neq 0$ in the classical Bonferroni case $\gamma = 1$.

In order to illustrate the PTS principle in relation with the classical Gini and Bonferroni welfare functions, corresponding to the extreme values of the parameter $\gamma = 0, 1$, consider first the classical Gini welfare function $A_G(\boldsymbol{x})$, whose weighting structure for $n = 8$ is given by (11) as follows,

$$\boldsymbol{w}^G = \left(\frac{15}{64}, \frac{13}{64}, \frac{11}{64}, \frac{9}{64}, \frac{7}{64}, \frac{5}{64}, \frac{3}{64}, \frac{1}{64}\right). \tag{36}$$

Consider now a progressive income transfer δ from $x_{(j)}$ to $x_{(i)}$ with $j = i + 1$. This transfer results in a new income distribution in which $x'_{(i)} = x_{(i)} + \delta$, $x'_{(j)} = x_{(j)} - \delta$, and $x'_{(k)} = x_{(k)}$ for $k \neq i, j$. According to the expression for the classical Gini welfare difference (22), we obtain

$$\begin{aligned}
&\text{for } i = 1,\ \ j = 2\ :\ A_G(\boldsymbol{x}') - A_G(\boldsymbol{x}) = (w_1^G - w_2^G)\delta = \frac{1}{32}\,\delta \\
&\text{for } i = 2,\ \ j = 3\ :\ A_G(\boldsymbol{x}') - A_G(\boldsymbol{x}) = (w_2^G - w_3^G)\delta = \frac{1}{32}\,\delta \\
&\text{for } i = 3,\ \ j = 4\ :\ A_G(\boldsymbol{x}') - A_G(\boldsymbol{x}) = (w_3^G - w_4^G)\delta = \frac{1}{32}\,\delta \\
&\text{for } i = 4,\ \ j = 5\ :\ A_G(\boldsymbol{x}') - A_G(\boldsymbol{x}) = (w_4^G - w_5^G)\delta = \frac{1}{32}\,\delta \\
&\text{for } i = 5,\ \ j = 6\ :\ A_G(\boldsymbol{x}') - A_G(\boldsymbol{x}) = (w_5^G - w_6^G)\delta = \frac{1}{32}\,\delta \\
&\text{for } i = 6,\ \ j = 7\ :\ A_G(\boldsymbol{x}') - A_G(\boldsymbol{x}) = (w_6^G - w_7^G)\delta = \frac{1}{32}\,\delta \\
&\text{for } i = 7,\ \ j = 8\ :\ A_G(\boldsymbol{x}') - A_G(\boldsymbol{x}) = (w_7^G - w_8^G)\delta = \frac{1}{32}\,\delta.
\end{aligned}$$

We can see that any progressive income transfer generates the same increase in welfare, meaning that the classical Gini welfare function does not satisfies PTS.

Consider now the classical Bonferroni welfare function $A_B(\boldsymbol{x})$, whose weighting structure for $n = 8$ is given by (15) as follows,

$$\boldsymbol{w}^B = \left(\frac{761}{2240}, \frac{481}{2240}, \frac{341}{2240}, \frac{743}{6720}, \frac{533}{6720}, \frac{73}{1344}, \frac{15}{448}, \frac{1}{64}\right). \tag{37}$$

As before, consider a progressive income transfer δ from $x_{(j)}$ to $x_{(i)}$ with $j = i+1$. This transfer results in a new income distribution in which $x'_{(i)} = x_{(i)} + \delta$, $x'_{(j)} = x_{(j)} - \delta$, and $x'_{(k)} = x_{(k)}$ for $k \neq i, j$. According to the expression for the classical Bonferroni welfare difference (23), we obtain

$$\begin{aligned}
&\text{for } i=1,\ \ j=2\ :\ A_B(\boldsymbol{x}')-A_B(\boldsymbol{x})=(w_1^B-w_2^B)\delta=\frac{1}{8}\,\delta\\
&\text{for } i=2,\ \ j=3\ :\ A_B(\boldsymbol{x}')-A_B(\boldsymbol{x})=(w_2^B-w_3^B)\delta=\frac{1}{16}\,\delta\\
&\text{for } i=3,\ \ j=4\ :\ A_B(\boldsymbol{x}')-A_B(\boldsymbol{x})=(w_3^B-w_4^B)\delta=\frac{1}{24}\,\delta\\
&\text{for } i=4,\ \ j=5\ :\ A_B(\boldsymbol{x}')-A_B(\boldsymbol{x})=(w_4^B-w_5^B)\delta=\frac{1}{32}\,\delta\\
&\text{for } i=5,\ \ j=6\ :\ A_B(\boldsymbol{x}')-A_B(\boldsymbol{x})=(w_5^B-w_6^B)\delta=\frac{1}{40}\,\delta\\
&\text{for } i=6,\ \ j=7\ :\ A_B(\boldsymbol{x}')-A_B(\boldsymbol{x})=(w_6^B-w_7^B)\delta=\frac{1}{48}\,\delta\\
&\text{for } i=7,\ \ j=8\ :\ A_B(\boldsymbol{x}')-A_B(\boldsymbol{x})=(w_7^B-w_8^B)\delta=\frac{1}{56}\,\delta.
\end{aligned}$$

We can see in this case that the actual position in which the progressive income transfer occurs has a differentiated impact on welfare. More specifically, the increase in welfare is greater when the transfer applies to the lowest income levels.

In general, we can measure the transfer sensitivity and positional transfer sensitivity of the GB welfare functions in terms of the parameter $\gamma \in [0,1]$ using the measures in (8) and (9) as follows,

$$TS(\gamma) = \sum_{i=1}^{n-1} w_i^{GB} - w_{i+1}^{GB} = w_1^{GB} - w_n^{GB}\,, \tag{38}$$

where w_i^{GB} are the weights of the single parameter GB welfare functions A_{GB} associated with the parameter $\gamma \in [0,1]$, with $i = 1,\dots,n$.

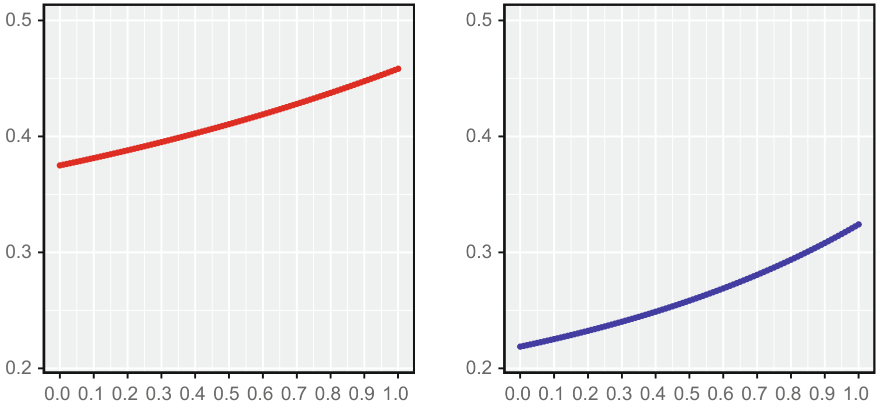
$$PTS(\gamma) = 1 + \sum_{i=1}^{n-1} \frac{\omega_i \ln \omega_i}{\ln(n-1)}\,, \tag{39}$$

where ω_i, with $i = 1,\dots,n-1$, is given by

$$\omega_i = \frac{w_i^{GB} - w_{i+1}^{GB}}{w_1^{GB} - w_n^{GB}}$$

where w_i^{GB} are the weights of the GB welfare functions, with $i = 1,\dots,n$.

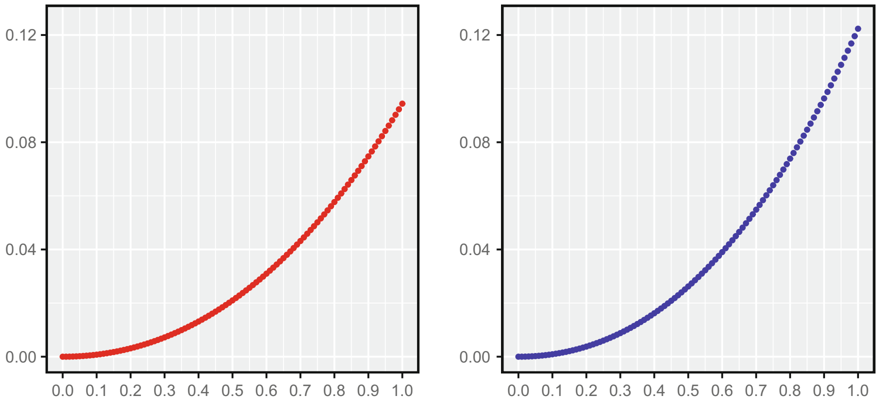
In Figs. 5 and 6 we can see the measures of transfer sensitivity and positional transfer sensitivity of the GB welfare functions associated with the parameter $\gamma \in [0,1]$, in the cases $n = 4, 8$. As the parameter γ value increases, we observe that both transfer sensitivity and positional transfer sensitivity of the A_{GB} welfare function increase too. Notice the fact that transfer sensitivity is not null for $\gamma = 0$, corresponding to the classical Gini case.



(a) TS in the case $n = 4$

(b) TS in the case $n = 8$

Fig. 5 Transfer sensitivity of the A_{GB} for parameter values $\gamma \in [0, 1]$, with $n = 4, 8$



(a) PTS in the case $n = 4$

(b) PTS in the case $n = 8$

Fig. 6 Positional transfer sensitivity of the A_{GB} for parameter values $\gamma \in [0, 1]$, with $n = 4, 8$

6 Conclusions

We have examined the binomial decomposition of the single parameter family of GB welfare functions and we have illustrated the dependence of the binomial decomposition coefficients in relation with the parameter which describes the GB family. We have found that $\alpha_1 = 1/n$ is independent of the parameter $\gamma \in [0, 1]$ and we have observed that only α_2 is decreasing, whereas $\alpha_3, \dots, \alpha_n$ are increasing with respect to $\gamma \in [0, 1]$. In particular, since the binomial coefficients α_j with $j = 1, \dots, n$ are non negative with unit sum, the decrease in α_2 compensates the increase in $\alpha_3, \dots, \alpha_n$

with respect to $\gamma \in [0, 1]$. The Bonferroni welfare function, associated with $\gamma = 1$, is obtained by means of this compensatory mechanism.

Moreover, we have illustrated the transfer sensitivity and positional transfer sensitivity of the binomial welfare functions, and we have examined their properties with respect to these principles. For this purpose, we have introduced measures of transfer sensitivity and positional transfer sensitivity for generalized Gini welfare functions and, in particular, we have illustrated the behaviour of these measures in the case of the GB welfare functions, in relation with the values of the parameter $\gamma \in [0, 1]$.

References

1. Aaberge, R.: Characterizations of Lorenz curves and income distributions. Soc. Choice Welf. **17**(4), 639–653 (2000)
2. Aaberge, R.: Axiomatic characterization of the Gini coefficient and Lorenz curve orderings. J. Econ. Theory **101**(1), 115–132 (2001)
3. Aaberge, R.: Ginis nuclear family. J. Econ. Inequality **5**(3), 305–322 (2007)
4. Aristondo, O., García-Lapresta, J.L., de la Vega, C.L., Pereira, R.A.M.: Classical inequality indices, welfare and illfare functions, and the dual decomposition. Fuzzy Sets Syst. **228**, 114–136 (2013)
5. Atkinson, A.B.: On the measurement of inequality. J. Econ. Theory **2**(3), 244–263 (1970)
6. Bárcena Martín, E., Imedio, L.J.: The Bonferroni, Gini, and De Vergottini indices. Inequality, welfare, and deprivation in the European union in 2000. In: Bishop, J., Zheng, B. (eds.) Inequality and Opportunity: Papers from the Second ECINEQ Society Meeting, Research on Economic Inequality, vol. 16, pp. 231–257. Emerald Group Publishing Limited (2008)
7. Bárcena Martín, E., Silber, J.: On the concepts of Bonferroni segregation index and curve. Rivista Italiana di Economia Demografia e. Statistica **65**(2), 57–76 (2011)
8. Bárcena Martín, E., Silber, J.: On the generalization and decomposition of the Bonferroni index. Soc. Choice Welf. **41**(4), 763–787 (2013)
9. Beliakov, G., Bustince Sola, H., Calvo, T.: A practical guide to averaging functions. Studies in Fuzziness and Soft Computing, vol. 329. Springer, Heidelberg (2016)
10. Beliakov, G., Pradera, A., Calvo, T.: Aggregation functions: a guide for practitioners. Studies in Fuzziness and Soft Computing, vol. 221. Springer, Heidelberg (2007)
11. Porath, E.B., Gilboa, I.: Linear measures, the Gini index, and the income-equality trade-off. J. Econ. Theory **64**, 443–467 (1994)
12. Benedetti, C.: Sulla interpretazione benesseriale di noti indici di concentrazione e di altri. Metron **44**(1), 421–429 (1986)
13. Blackorby, C., Donaldson, D.: Measures of relative equality and their meaning in terms of social welfare. J. Econ. Theory **18**(1), 59–80 (1978)
14. Blackorby, C., Donaldson, D.: Ethical indices for the measurement of poverty. Econometrica **48**(4), 1053–1060 (1980)
15. Blackorby, C., Donaldson, D.: A theoretical treatment of indices of absolute inequality. Int. Econ. Rev. **21**(1), 107–136 (1980)
16. Blackorby, C., Donaldson, D.: Ethical social index numbers and the measurement of effective tax/benefit progressivity. Can. J. Econ. **17**(4), 683–694 (1984)
17. Blackorby, C., Donaldson, D., Auersperg, M.: A new procedure for the measurement of inequality within and among population subgroups. Can. J. Econ. **14**(4), 665–685 (1981)
18. Bonferroni, C.E.: Elementi di Statistica Generale. Seeber, Firenze (1930)
19. Bonferroni, C.E.: Elementi di Statistica Generale. Ristampa con aggiunte, anno accademico 1932–1933. Litografia Felice Gili, Torino (1933)

20. Bortot, S., Pereira, R.A.M.: The binomial Gini inequality indices and the binomial decomposition of welfare functions. Fuzzy Sets Syst. **255**, 92–114 (2014)
21. Bossert, W.: An axiomatization of the single-series Ginis. J. Econ. Theory **50**(1), 82–92 (1990)
22. Calvo, T., De Baets, B.: Aggregation operators defined by k-order additive/maxitive fuzzy measures. Int. J. Uncertainty Fuzziness Knowl.-Based Syst. **6**(6), 533–550 (1998)
23. Calvo, T., Kolesárová, A., Komorníková, M., Mesiar, R.: Aggregation operators: properties, classes and construction methods. In: Calvo, T., Mayor, M., Mesiar, R. (eds.) Aggregation Operators: New Trends and Applications. Studies in Fuzziness and Soft Computing, vol. 97, pp. 3–104. Springer, Heidelberg (2002)
24. Chakravarty, S.R.: The Bonferroni indices of inequality. In: International Conference in Memory of C. Gini and M. O. Lorenz (2005)
25. Chakravarty, S.R.: A deprivation-based axiomatic characterization of the absolute Bonferroni index of inequality. J. Econ. Inequality **5**(3), 339–351 (2007)
26. Chakravarty, S.R., Muliere, P.: Welfare indicators: a review and new perspectives. measurement of inequality. Metron **61**(3), 457–497 (2003)
27. Chisini, O.: Sul concetto di media. Periodico di matematiche **9**(4), 106–116 (1929)
28. De Vergottini, M.: Sul significato di alcuni indici di concentrazione. Giornale degli Economisti e Annali di Economia **2**(5/6), 317–347 (1940)
29. De Vergottini, M.: Sugli indici di concentrazione. Statistica **10**(4), 445–454 (1950)
30. Donaldson, D., Weymark, J.A.: A single-parameter generalization of the Gini indices of inequality. J. Econ. Theory **22**(1), 67–86 (1980)
31. Donaldson, D., Weymark, J.A.: Ethically flexible Gini indices for income distributions in the continuum. J. Econ. Theory **29**(2), 353–358 (1983)
32. Ebert, U.: Size and distribution of incomes as determinants of social welfare. J. Econ. Theory **41**(1), 23–33 (1987)
33. Ebert, U.: Measurement of inequality: an attempt at unification and generalization. Soc. Choice Welf. **5**(2), 147–169 (1988)
34. Fodor, J., Marichal, J.L., Roubens, M.: Characterization of the ordered weighted averaging operators. IEEE Trans. Fuzzy Syst. **3**(2), 236–240 (1995)
35. Fodor, J., Roubens, M.: Fuzzy preference modelling and multicriteria decision support. Theory and Decision Library, series D, vol. 14. Kluwer Academic Publishers, Dordrecht (1994)
36. Gajdos, T.: Measuring inequalities without linearity in envy: choquet integrals for symmetric capacities. J. Econ. Theory **106**(1), 190–200 (2002)
37. Gini, C.: Variabilità e mutabilità Contributo alla studio delle Distribuzioni e delle Relazioni Statistiche. P. Cuppini, Bologna (1912)
38. Gini, C.: Sulla misura della concentrazione e della variabilità dei caratteri. Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti **78**, 1203–1248 (1914) (English translation given in Metron **63**, 1–38 (2005))
39. Gini, C.: Measurement of inequality of incomes. Econ. J. **31**(121), 124–126 (1921)
40. Giorgi, G.M.: A methodological survey of recent studies for the measurement of inequality of economic welfare carried out by some Italian statisticians. Econ. Notes **13**(1), 146–157 (1984)
41. Giorgi, G.M.: Concentration index, Bonferroni. In: Kotz, S., Read, C.B., Banks, D.L. (eds.) Encyclopedia of Statistical Sciences, pp. 141–146. Wiley, New York (1998)
42. Giorgi, G.M., Crescenzi, M.: A look at the Bonferroni inequality measure in a reliability framework. Statistica **61**(4), 571–584 (2001)
43. Giorgi, G.M., Crescenzi, M.: A proposal of poverty measures based on the Bonferroni inequality index. Metron **59**(3–4), 3–16 (2001)
44. Giorgi, G.M., Nadarajah, S.: Bonferroni and Gini indices for various parametric families of distributions. Metron **68**(1), 23–46 (2010)
45. Grabisch, M.: k-order additive discrete fuzzy measures and their representation. Fuzzy Sets Syst. **92**(2), 167–189 (1997)
46. Grabisch, M., Marichal, J.L., Mesiar, R., Pap, E.: Aggregation functions. Encyclopedia of Mathematics and its Applications, vol. 127. Cambridge University Press (2009)

47. Kolm, S.C.: The optimal production of social justice. In: Margolis, J., Guitton, H. (eds.) Public Economics. International Economic Association Series, pp. 145–200. Macmillan, London (1969)
48. Kolm, S.C.: Unequal inequalities I. J. Econ. Theory **12**(3), 416–442 (1976)
49. Kolm, S.C.: Unequal inequalities II. J. Econ. Theory **13**(1), 82–111 (1976)
50. Lorenzen, G.: A Generalized Gini-Coefficient. Universität Hamburg, Institut für Statistik und Ökometrie (1979)
51. Marshall, A.W., Olkin, I.: Inequalities: Theory of Majorization and Its Applications. Mathematics in Science and Engineering, vol. 143. Academic Press, New York (1979)
52. Mehran, F.: Linear measures of income inequality. Econometrica **44**(4), 805–809 (1976)
53. Nygård, F., Sandström, A.: Measuring Income Inequality. Almqvist & Wiksell International, Stockholm (1981)
54. Quiggin, J. (ed.): Generalized Expected Utility Analysis: The Rank-Dependent Model. Kluwer Academic Publishers, Dordrecht (1993)
55. Sen, A.: On Economic Inequality. Claredon Press, Oxford (1973)
56. Sen, A.: Ethical measurement of inequality: some difficulties. In: Krelle, W., Shorrocks, A.F. (eds.) Personal Income Distribution. North-Holland, Amsterdam (1978)
57. Shorrocks, A.F., Foster, J.E.: Transfer sensitive inequality measures. Rev. Econ. Stud. **54**(3), 485–497 (1987)
58. Silber, J.: Factor components, population subgroups and the computation of the Gini index of inequality. Rev. Econ. Stat. **71**(1), 107–115 (1989)
59. Tarsitano, A.: The Bonferroni index of income inequality. In: Dargum, C., Zenga, M. (eds.) Income and Wealth Distribution. Inequality and Poverty, pp. 228–242. Springer, Heidelberg (1990)
60. Weymark, J.A.: Generalized Gini inequality indices. Math. Soc. Sci. **1**(4), 409–430 (1981)
61. Yaari, M.E.: The dual theory of choice under risk. Econometrica **55**(1), 95–115 (1987)
62. Yaari, M.E.: A controversial proposal concerning inequality measurement. J. Econ. Theory **44**(2), 381–397 (1988)
63. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Trans. Syst. Man Cybern. **18**(1), 183–190 (1988)
64. Yager, R.R., Kacprzyk, J.: The Ordered Weighted Averaging Operators: Theory and Applications. Kluwer Academic Publisher, Dordrecht (1997)
65. Yager, R.R., Kacprzyk, J., Beliakov, G.: Recent developments in the ordered weighted averaging operators: theory and practice. Studies in Fuzziness and Soft Computing, vol. 265. Springer, Heidelberg (2011)
66. Yitzhaki, S.: On an extension of the Gini inequality index. Int. Econ. Rev. **24**(3), 617–628 (1983)
67. Zenga, M.M.: Decomposition by sources of the Gini, Bonferroni and Zenga inequality indexes. Stat. Appl. **11**(2), 133–161 (2013)
68. Zoli, C.: Intersecting generalized Lorenz curves and the Gini index. Soc. Choice Welf. **16**(2), 183–196 (1999)