Handling a Large Number of Preferences in a Multi-Level Decision-Making Process

Ana Teresa Tapia Rosero

Promotor: prof. dr. G. De Tré
Proefschrift ingediend tot het behalen van de graad van
Doctor in de ingenieurswetenschappen

Vakgroep Telecommunicatie en Informatieverwerking
Voorzitter: prof. dr. ir. H. Bruneel
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2015 - 2016

UNIVERSITEIT GENT

ISBN 978-90-8578-914-7
NUR 982, 984
Wettelijk depot: D/2016/10.500/46

Members of the Examination Board

Prof. dr. ir. Patrick De Baets	(Ghent University, Chair)
Prof. dr. ir. Tom Dhaene	(Ghent University, Secretary)
Prof. dr. Luis Martínez López	(University of Jaén)
Prof. dr. Enrique Peláez Jarrín	(ESPOL University)
Prof. dr. Rosa Rodríguez Domínguez	(University of Granada)
Prof. dr. ir. Sidharta Gautama	(Ghent University)
Prof. dr. Guy De Tré	(Ghent University, Supervisor)

Affiliations

Research Group for Database Document and Content Management (DDCM)
Department of Telecommunications and Information Processing (TELIN)
Faculty of Engineering and Architecture
Ghent University
Sint-Pietersnieuwstraat 41
B-9000 Ghent
Belgium

Facultad de Ingeniería en Electricidad y Computación (FIEC)
Escuela Superior Politécnica del Litoral (ESPOL)
Campus Gustavo Galindo V.
Km. 30.5 via Perimetral
Guayaquil - Ecuador

To “mi Amorcito” and “mi Principito”
for being the most important part
of this amazing journey of adventures.

I love you so much!
♡

Samenvatting

Het nemen van beslissingen is een proces dat elk van ons dagelijks ervaart. Dat varieert van eenvoudige beslissingen die met gemak worden genomen tot complexere beslissingen waarvoor intensiever nadenken is vereist. Bij sommige beslissingen kunnen verscheidene personen betrokken zijn, waaronder meerdere beslissingsnemers. Bijvoorbeeld, afhankelijk van het beslissingsprobleem, kunnen bij een beslissingsproces aan een academische instelling zowel studenten, professoren, administratieve medewerkers als academische beleidsvoerders betrokken zijn. In dit voorbeeld lijken de deelnemers aan het beslissingsproces één van de hoofdoorzaken van de complexiteit van het beslissingsprobleem te zijn.

Zonder verlies van algemeenheid, kunnen er bij een beslissingsproces twee types van deelnemers worden geïdentificeerd: de reguliere deelnemers en de beslissingsnemers. Een reguliere deelnemer is iemand die betrokken is in het proces door zijn/haar voorkeuren uit te drukken met betrekking tot de beslissingscriteria, zonder daarbij te beslissen wat het resultaat is. Een beslissingsnemer is een persoon die zijn/haar perspectief (bvb. sociaal, technisch, financieel of milieu) kenbaar maakt en beslist over het resultaat van een beslissingsprobleem.

Naast het type deelnemers, kan de complexiteit van een beslissingsprobleem toenemen door een groot aantal van hen. Bovendien kan de complexiteit nog verder worden verhoogd wanneer wordt verondersteld dat de voorkeuren van sommige deelnemers belangrijker zijn dan deze van anderen. Hoewel de complexiteit is toegenomen, kan een groot aantal deelnemers, samen met de diversiteit aan preferenties (gegeven door de reguliere deelnemers) en meerdere perspectieven (gegeven door de beslissingsnemers) bijdragen tot beter gemotiveerde beslissingen. Daarom wordt het probleem van het behandelen van de complexiteit van het nemen van beslissingen in de aanwezigheid van een groot aantal preferenties en meerdere perspectieven bestudeerd in dit proefschrift. Daarbij worden drie onderzoeksvragen, in de context van het nemen van beslissingen, behandeld: (i) Hoe kan men een groot aantal preferenties behandelen? (ii) Hoe kan men preferenties, die relevant zijn vanuit het perspectief van een beslissingsnemer, identificeren en evalueren? En (iii) Hoe kunnen preferenties die afkomstig zijn uit meerdere perspectieven worden gecombineerd?

In het kader van de eerste onderzoeksvraag over het behandelen van een groot aantal preferenties, wordt een nieuwe 'vormgelijkenis'-detectiemethode

voorgesteld. Deze methode beoogt om de complexiteit van het probleem van het nemen van een beslissing, dat voortvloeit uit het aantal preferenties dat gegeven wordt door reguliere deelnemers, te reduceren. Daartoe groepeert de methode de op elkaar gelijkende preferenties, zodat de beslissingsnemer niet meer moet omgaan met individuele preferenties, maar met een gereduceerde verzameling van gegroepeerde, op elkaar gelijkende preferenties. Om de op elkaar gelijkende preferenties te groeperen, gebruikt de voorgestelde methode een voorstelling waarbij de preferenties van de deelnemers worden gemodelleerd als lidmaatschapsfuncties (d.i., door gebruik te maken van preferentiemodellering met vaagverzamelingen). Deze lidmaatschapsfuncties worden dan geannoteerd met een nieuwe 'vorm-symbolische' notatie. Deze 'vorm-symbolische' notatie heeft twee componenten, namelijk een vorm-string en een kenmerk-string. De vorm-string duidt de vormkarakteristieken van de lidmaatschapsfunctie aan, zoals de schuinten en niveaus van preferentie, terwijl de kenmerk-string de relatieve lengte van de vormkarakteristieken op hun X-as segmenten uitdrukt, d.i. de kernsegmenten en de linker- en rechterschuinten. Bijkomend wordt ook een originele 'vormgelijkenis'-maat voorzien om 'vorm-symbolische' annotaties van lidmaatschapsfuncties te kunnen vergelijken. Deze maat wordt dan gebruikt in een clusteringsproces om een groot aantal preferenties te partitioneren in een gereduceerd aantal groepen van op elkaar gelijkende preferenties.

Om de tweede onderzoeksvraag over het identificeren en evalueren van groepen van preferenties, die als relevant worden beschouwd volgens het perspectief van één enkele beslissingsnemer te behandelen, wordt een nieuwe methodologie voorgesteld. Deze methodologie bevat: (i) een model om preferenties te aggregeren over groepsattributen die het perspectief van de beslissingsnemer weergeven, (ii) een cohesiemaat als karakteristiek (of attribuut) van groepen die werden aangemaakt via de 'vormgelijkenis'-detectiemethode, (iii) twee berekeningsmethodes om de cohesiemaat van een groep te bepalen, en (iv) een procedure om de relevantie van een groep te berekenen op basis van het voorziene aggregatiemodel. Het aggregatiemodel (over groepsattributen) houdt rekening met het feit dat reguliere deelnemers verschillende achtergronden kunnen hebben, d.i., verschillende opleidingsniveaus, expertisegebieden en persoonlijke profielen. Binnen het model worden drie attributen gebruikt om groepen van preferenties te identificeren die vanuit het perspectief van de beslissingsnemer als relevant worden beschouwd. Deze attributen zijn 'groepsgrootte', 'aantal opmerkelijke preferenties (of opinies)' en 'cohesie'. Terwijl het attribuut 'groepsgrootte' het aantal lidmaatschapsfuncties dat bevat is in de groep voorstelt, duidt het attribuut 'aantal opmerkelijke preferenties' aan hoeveel opinies uit de groep enige extra aandacht waard zijn (of belangrijker zijn); en correspondeert het attribuut 'cohesie' met een maat die het niveau van vertrouwen uitdrukt in een groep die gevormd wordt door lidmaatschapsfuncties met een gelijksoortige vorm. De cohesie is een maat voor de overeenstemming tussen de lidmaatschapsfuncties die deel uitmaken van een groep (of cluster) die gegroepeerd werd op basis van vormgelijkenis, waarbij een hogere waarde een groter vertrouwen in de groep aanduidt. In dit proefschrift worden twee nieuwe benaderingen om een cohesiemaat te berekenen bestudeerd, namelijk een berekening door middel van een uitgebreide 'vorm-symbolische' notatie en een berekening die uitgaat van een geometrische benadering. De idee achter beide benaderingen is dat, voor de berekening van een cohesiemaat een groep van lidmaatschapsfuncties van preferenties kan worden gekarakteriseerd door een boven- en ondergrens. De karakterisering binnen de benadering met een uitgebreide 'vorm-symbolische' notatie gebruikt een triplet van de vorm ⟨vorm-string, kenmerk-string, breedte-string⟩. In dit geval wordt de cohesiemaat berekend door rekening te houden met de numerieke waarden die geassocieerd zijn met de componenten van de uitgebreide 'vorm-symbolische' annotatie. Bij een geometrische benadering wordt de cohesiemaat berekend op basis van de oppervlakte die bevat is tussen de boven- en ondergrens van de groep van lidmaatschapsfuncties. Om beide benaderingen te vergelijken werden drie verschillende scenario's bestudeerd. Het eerste scenario correspondeert met het geval waarbij alle preferenties door dezelfde lidmaatschapsfunctie worden voorgesteld. Het tweede scenario is een typisch geval waarbij verscheidene opinies in grote mate overeenkomen en dus ook worden voorgesteld door lidmaatschapsfuncties met een gelijksoortige vorm. Terwijl het derde scenario een atypisch geval voorstelt, waarbij een uitschieter voorkomt in een groep van overeenstemmende preferenties. In de studie wordt aangetoond dat beide benaderingen de verwachte resultaten weergeven. Daarnaast wordt aangetoond dat een procedure om de relevantie van een groep te berekenen door de hiervoor vermelde groepsattributen (d.i., groepsgrootte, aantal opmerkelijke preferenties en cohesie) te combineren, het mogelijk maakt om een indicator voor de relevantie van een groep van preferenties te bouwen. Eens een groep van preferenties kan worden gevalueerd vanuit het perspectief van één enkele beslissingsnemer, is het mogelijk om meerdere perspectieven te behandelen.

Om de derde onderzoeksvraag over het combineren van preferenties uit meerdere perspectieven aan te pakken, wordt een nieuw fusiemodel voor beslissingsondersteuning voorgesteld. Dit model is gebaseerd op een abstracte voorstelling die een decision-making unit of kort DMU wordt genoemd. Een DMU laat het toe om preferenties die als input worden verkregen te fusioneren met de preferenties die worden opgegeven door één enkele beslissingsnemer. De preferenties van een beslissingsnemer weerspiegelen zijn/haar perspectief op het beslissingsprobleem, terwijl de preferenties die als input worden verkregen afkomstig kunnen zijn van een groot aantal personen die betrokken zijn bij de beslissing (d.i., reguliere deelnemers) of van andere DMU's in een hiërar-chische structuur. Een hiërarchische structuur met DMU knopen wordt gebruikt om de organisatorische structuur van het beslissingsmodel voor te stellen en om de propagatie van preferenties te vergemakkelijken. Hierbij geeft de organisatorische structuur algemeen aan hoe meerdere beslissingsnemers participeren in een beslissingsproces. In dit proefschrift bestaat de propagatie van preferenties uit het verrijken van de informatie (op het niveau) waar de finale beslissing wordt genomen. Daardoor kan een beslissingsnemer op het hoogste organisatorisch niveau een beter gemotiveerde beslissing nemen.

In haar totaliteit draagt dit proefschrift bij tot de studie van beslissingsproblemen waarbij verschillende kennisdomeinen, d.i., verschillende perspectieven gegeven door personen met verschillende expertisegebieden, in rekening dienen te worden gebracht en waarbij het mogelijk is om de deeltaken van een beslissing te delegeren. Bijvoorbeeld, een beslissingsprobleem in een multinationale onderneming met afdelingen in meerdere landen, waarbij de hoofdzetel rekening wenst te houden met de meningen van regionale (en sub-regionale) organisatorische eenheden en hun respectievelijke klanten. In dit voorbeeld kan elke regionale (of sub-regionale) manager regionale beperkingen die gerelateerd zijn aan zijn/haar competentiegebied (bvb. cultuur, milieu, financieel) toevoegen aan zijn/haar perspectief.

Summary

Decision making is a process that all of us experience daily, from habitual decisions which are made with ease to more complex ones where more intensive thinking is needed. Moreover, a decision-making process can involve several persons including multiple decision makers. For example, depending on the decision problem, a decision process within an academic institution can involve its students, professors, administrative staff as well as the academic authorities. In the example, the participants in the decision process seem to be one of the keys regarding the complexity of the decision-making problem.

Without loss of generality, two types of participants can be identified in a decision-making process: the regular participants and the decision makers. A regular participant is a person who is involved in the process by expressing his/her preferences regarding the decision criteria, but without deciding what the result is. A decision maker is a person who expresses his/her perspective (e.g., social, technical, financial or environmental) and decides on the result of a decision problem.

Besides the type of participants, the complexity of a decision-making problem can be increased by a large number of them. Moreover, the complexity may be further increased when it is considered that the preferences given by some participants are more relevant than others. Although the complexity is increased, a large number of participants together with the diversity of preferences (given by the regular participants) and multiple perspectives (given by the decisions makers) can contribute to make better motivated decisions. Therefore, the problem of handling the complexity of decision making in the presence of a large number of preferences and multiple perspectives is studied in this dissertation. Herein, three main research questions are addressed in a decision-making context: (i) How to handle a large number of preferences? (ii) How to identify and evaluate preferences considered being relevant from a decision maker's perspective? And (iii) How to combine preferences from multiple perspectives?

To cope with the first research question of handling a large number of preferences, a novel shape-similarity detection method is proposed. This method aims to reduce the complexity of a decision-making problem related to the number of preferences given by regular participants. For this purpose, the method groups similar preferences in such a way that a decision maker does not have to deal with all the individual preferences anymore, but with a reduced set of grouped

similar preferences. To group similar preferences, the proposed method uses the preferences of the participants represented as membership functions (i.e., using preference modeling with fuzzy sets). Those membership functions are then annotated using a novel shape-symbolic notation. This shape-symbolic notation has two components, namely a shape-string and a feature-string. The shape-string component denotes the shape characteristics of the membership function such as slopes and levels of preference, while the feature-string component expresses the relative length of the shape characteristics on their X-axis segments, i.e. the core segments and the left and right spreads. Additionally, to facilitate the comparisons among shape-symbolic annotations of membership functions, an original shape-similarity measure is provided. This measure is then used within a clustering process to partition a large amount of preferences into a reduced amount of groups of similar preferences.

To handle the second research question of identifying and evaluating groups of preferences considered to be relevant according to a single decision maker's perspective, a novel methodology is provided. This methodology includes: (i) a model for aggregating preferences on group attributes reflecting a decision maker's perspective, (ii) a cohesion measure as a characteristic (or attribute) of groups formed by the shape-similarity detection method, (iii) two computational methods to obtain the cohesion measure of a group, and (iv) a procedure to compute the relevance of a group based on the provided aggregation model. The aggregation model (on group attributes) takes into account that regular participants have different backgrounds, i.e., different education levels, areas of expertise and personal profiles. Within this model, three attributes are used to identify groups of preferences that are considered to be relevant from a decision maker's perspective. These attributes are 'group size', 'number of noticeable preferences (or opinions)' and 'cohesion'. While the attribute 'group size' represents the amount of membership functions contained in the group, the attribute 'number of noticeable preferences' denotes the number of preferences in the group that are worthy some extra attention (or are more important); and, the attribute 'cohesion' corresponds to a measure denoting the level of confidence in a group that is formed by similarly shaped membership functions. The cohesion is a measure for the level of togetherness among membership functions that are part of a group (or cluster) grouped by shape-similarity where a higher value indicates more confidence in the group. In this dissertation two novel approaches to compute a cohesion measure are studied, namely a computation by means of an extended shape-symbolic notation and a computation that departs from a geometric approach. The idea behind both approaches is that, to compute the cohesion measure, a group of membership functions of preferences can be characterized by an upper and a lower bound. The characterization within an extended shape-symbolic notation approach uses a triplet having the form ⟨shape-string, feature-string, width-string⟩. In this case, the cohesion measure is computed taking into account the numerical values associated to the components of the extended shape-symbolic annotation. Within the geometric approach, the cohesion measure is computed taking into account the area contained within the upper and lower bounds of the group of membership functions. To compare both approaches three different scenarios have been studied. The first scenario corresponds to the case where all preferences are represented by the same membership function. The second scenario is a typical case where several opinions mostly agree and hence are represented by similarly shaped membership functions. While the third scenario represents an atypical case in which an outlier exists in a group of similar preferences. It is shown, herein, that both approaches reflect the expected results. Beside of that, it is shown that a procedure to compute the relevance of a group by combining the aforementioned group attributes (i.e., group size, number of noticeable opinions and cohesion) makes it possible to obtain an indicator of the relevance of a group of preferences. Once a group of preferences can be evaluated from a single decision maker's perspective, it is possible to handle multiple perspectives.

To address the third research question of combining preferences from multiple perspectives, a novel fusion model for decision support is proposed. This model is based on an abstract representation called decision-making unit or DMU for short. A DMU allows for the fusion of preferences received as inputs with the preferences provided by a single decision maker. The decision maker's preferences reflect his/her perspective of the decision problem, while the preferences received as inputs might come from a large number of persons involved in the decision (i.e., regular participants), as well as other DMUs in a hierarchical structure. A hierarchical structure with DMU nodes is used to represent an organizational structure of the decision model and to facilitate the propagation of preferences. Herein, the organizational structure generally reflects how multiple decision makers participate in a decision-making process. In this dissertation, the propagation of preferences consists in enriching the information (at the level) where a final decision is made. In this way, the decision maker at the highest organizational level can make a better motivated decision.

Overall this dissertation contributes to the study of decision problems that involve different domains of knowledge, i.e. different perspectives given by persons with different areas of expertise, where it is possible to delegate the sub-tasks of decision making. For example, a decision-making problem in a multinational corporation with operations in more than one country, where the headquarters would like to take into account the opinions given by the regional (and sub-regional) organizational units and their respective customers. In this example, each regional (or sub-regional) manager may include in his/her perspective the regional constraints (e.g., cultural, environmental, financial, among others) that are related to his/her competence area.

Acknowledgements

"Feeling gratitude and not expressing it
is like wrapping a present and not giving it."
—William Arthur Ward.

I would like to thank every person that has contributed to make the journey of my doctoral studies a lifetime experience, but the space in this dissertation might not be enough. So, only a few gratitude words are expressed next.

I would like to express my gratitude to God for all the blessings and opportunities that He puts everyday in my way.

To my adviser, Guy, for his wise guidance and help on so many aspects. Thank you for making our stay such a fruitful, pleasant and memorable experience. I am also thankful to your family for all the shared time and for making us to feel at home.

Special thanks to the DDCM team for all the help and bright ideas that improved my work, along with the jokes, laughs and enjoyable moments. Thank you all for being so kind to me and my family. Antoon, Christophe, Robin, Joachim and Sofian: thanks from the deepest of my heart! To the TELIN family thanks for all the hospitality, support, good times, and so much more. I really appreciate it all.

I am thankful to Katherine, Carlos, Xavier, Mónica and Edgar for their guidance since I started to look for a foreign university and a research group. To Guido for all the training and the good vibes, thanks! My gratitude to Prof. Sergio Flores for his wise advices and his support. To his lovely wife for making us feel as part of the family. ¡Gracias de todo corazón!

To Escuela Superior Politécnica del Litoral (ESPOL) and Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) for the financial support during my doctoral studies.

I would like to express my appreciation to each of the members of the Examination Board for their valuable comments and suggestions that led to an improved version of this dissertation.

I would like to thank my family. My mom, dad, sister and my godparents for being by my side in spite of the distance. I love you! Gracias mami por ser mi mujer modelo, tú eres una verdadera guerrera de la luz. Papi gracias por mil y un cosas, entre esas el "obligarme" a aprender inglés durante mis vacaciones, las siguientes páginas solo muestran una parte del resultado. Ustedes siempre me han mostrado de la mejor forma posible la importancia de estudiar, ser puntual y hacer las cosas bien: lo hicieron con el ejemplo. ¡Los quiero un montón! Ñañi, gracias por los tips y por de vez en cuando distraerme sanamente. Me encantaron todas las sorpresas que nos enviaste sin ser Navidad. Gracias por todos esos detalles, fueron muchos y muy valiosos. Eres mi ñañi favorita ;) Tía Lupe, gracias por muchas cosas entre esas por estar pendiente de la familia y estar dispuesta a ayudar siempre. ¡Tienes un corazón de oro! Tío Daniel, gracias por enseñarme que el cielo es el límite y gracias por todos esos escalones que me has ayudado a poner. No hay palabras que me permitan expresarte todo el agradecimiento que te tengo ¡Te quiero un mundo!

Thanks to the Family Delcloo for their friendship and support on a variety of activities. Inge, Andy, Matthias, Lucas and Hanne: thanks for all the shared times! That is priceless! I will never forget how Lucas helped Marcelo when he barely spoke the language. Moreover, how Hanne smiles at us and how Matthias waves or stops his bike to say 'hi!' anytime and anywhere when they see us. We hope that the next ping-pong game between our families takes place in Ecuador, until then we'll be practicing :)

I will be forever thankful to my husband and my son for their company, help, support and patience. This adventure would not have been possible without you. ¡Gracias por hacerme barra! Amorcito gracias por ayudarme a ordenar mis ideas y mi eterno agradecimiento por ayudarme a ser una versión mejorada de mí misma. Marce Bubú gracias por ser tan cariñoso y comprensivo. Gracias por darme cada día ese maravilloso momento que me llena de energía. Estoy segura que tú sabes a cuál me refiero... ¡sí, ese mismo! Este es el libro que te conté, ese que estaba escribiendo y que veías a papi revisar. Aunque este no es uno de esos libros que pueda leerte antes de dormir, su publicación nos permitirá dedicar más tiempo a la lectura de otros libros que nos gusten a ambos.

Finally, thanks to my friends for providing the updates regarding different topics including special events such as weddings and births. Your greetings, emails, postcards, photos and chocolates made me feel closer to you.

Ana Tapia Rosero
Ghent 2016.

Contents

Acronyms

AHP	Analytic Hierarchy Process
DM	Decision Maker
DMP	Decision-Making Process
DMU	Decision Making Unit
GCD	Generalized Conjunction/Disjunction
GDM	Group Decision Making
IVFS	Interval-Valued Fuzzy Set
LSP	Logic Scoring of Preferences
MCDM	Multicriteria Decision-Making
MCGDM	Multicriteria Group Decision-Making
OWA	Ordered Weighted Average
TOPSIS	Technique for the Order Preference by Similarity to Ideal Solution
VIP	Very Important Person
WPM	Weighted Power Means

Publications

Publications

Articles in journals included in the Science Citation Index, Social Science Citation Index or Arts and Humanities Citation Index (A1)

1. Tapia-Rosero Ana, Bronselaer Antoon and De Tré Guy. A method based on shape-similarity for detecting similar opinions in group decision making. *Information Sciences*, 258 (2014): 291-311.

2. Tapia-Rosero Ana, Bronselaer Antoon, De Mol Robin and De Tré Guy. Fusion of preferences from different perspectives in a decision-making context. *Information Fusion*, 29 (2016): 120-131.

Articles in proceedings of scientific conferences, included in the Science Citation Index, Social Science Citation Index and Arts and Humanities Citation Index (P1)

1. De Mol Robin, Tapia-Rosero Ana and De Tré Guy. An approach for uncertainty aggregation using generalised conjunction/disjunction aggregators. In *Proceedings of 16th World Congress of the International Fuzzy Systems Association (IFSA) and 9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)*, 1499-1506. Gijón, Asturias, Spain, 2015.

Articles in proceedings of scientific conferences, not included in previous sections (C1)

1. Tapia-Rosero Ana, Bronselaer Antoon and De Tré Guy. Similarity of membership functions: a shaped based approach. In *Proceedings of the 4th International Joint Conference on Computational Intelligence*, 402-409. Barcelona, Spain, 2012.

2. Tapia-Rosero Ana and De Tré Guy. Evaluating relevant opinions within a large group. In *Proceedings of the 6th International Joint Conference on Computational Intelligence*, 402-409. Rome, Italy, 2014.

Articles included as book chapters, not included in previous sections (B2)

1. Tapia-Rosero Ana and De Tré Guy. A cohesion measure for expert preferences in group decision-making. In *Modern approaches in fuzzy sets, intuitionistic fuzzy sets, generalized nets and related topics. Volume II: Applications* edited by Krassimir Atanassov, Michal Baczynski, Jozef Drewniak, Janusz Kacprzyk, Maciej Krawczak, Eulalia Szmidt, Maciej Wygralak and Slawomir Zadrozny, 125-142, SRI-PAS, 2014.

2. Tapia-Rosero Ana, De Mol Robin and De Tré Guy. Handling uncertainty degrees in the evaluation of relevant opinions within a large group. In *Studies in Computational Intelligence (SCI)*, edited by Juan Julian Merelo, Agostinho Rosa, José M. Cadenas, António Dourado, Kurosh Madani and Joaquim Filipe, 283-299. Vol. 620, Springer International Publishing, 2016.

Award

- *Best Student Paper Award*[1] at the 6th International Conference on Fuzzy Computation Theory and Applications (FCTA 2014) as part of the 6th International Joint Conference on Computational Intelligence, for the paper *Evaluating relevant opinions within a large group.* Rome, Italy, October 24th, 2014.

[1] http://www.fcta.ijcci.org/PreviousAwards.aspx

List of Figures

List of Tables

Chapter 1

The Complexity of Decision-Making

The author starts this dissertation by explaining what is meant by the complexity of decision-making, whereby it facilitates describing the purpose of this research study. Next, some of the existing decision-making models regarding this research are briefly described, followed by the scope of this dissertation in the area of computational intelligence. Then, the research questions to address the challenges of this dissertation are stated. This chapter concludes describing the importance of this research study by describing how the existing decision-making models assist a decision maker managing the complexity in a decision and what is needed to address these days' challenges.

1.1 Introduction

Daily, all of us experience decisions with different levels of complexity: from habitual decisions which are made with ease, to more complex ones where more intensive thinking is needed. It is also possible that a decision considered to be simple at the beginning might become more complex by changing some aspects as illustrated in the following examples.

Choosing a film to watch may be a simple decision when someone goes alone to a particular cinema. Here, the decision is mainly based on the type of film that he/she can be interested in among the available films —e.g., based on the script, genre, cast, values, film reviews or comments given by friends. In the case of going accompanied by someone else, the complexity of this decision slightly increases since it depends on the type of film that both prefer and, moreover, the imprecision when expressing these preferences —e.g., while one might prefer films rated with at least $3^1/_2$ stars, the other might consider that films rated with at least 3 stars are acceptable, but films with 4 or more stars are preferred. Similarly, when someone is hosting a party at the cinema, the complexity of choosing a film increases even more considering the preferences of all of his/her guests. In this way, the complexity in the decision about "choosing a film to watch" increases when looking for a film that satisfies the group interests.

In this example, the following participants can be identified in the decision-making process: the guests and the host. Here, the guests are involved in the process by expressing their preferences regarding the films, but without deciding on the final selection of the film. The host is the person who makes the decision about which film to watch. In general, one may refer to the guests as (regular) *participants* and the host as a *decision maker*.

Choosing the sport facilities in a new park has some complexity when it is seen as a decision task for a city's Mayor and Councilors. In this case, the complexity of the decision is given by the presence of a more diverse group of interests where the city's Mayor and Councilors may have different personal profiles and different areas of expertise —e.g., engineering, public health, social welfare, environmental studies, among others. Additionally, decisions within an organization like a city hall should usually take into account some constraints that may affect the ability to carry out this work —e.g., financial resources or a time schedule. Here, the complexity of the decision may increase even more when it takes into account the preferences (e.g., which sports, the size for different facilities, the target group based on age and gender) of a community (e.g., the city residents), and the fact that some preferences can be worthy some extra attention or can be considered to be more important than others. In this way, the complexity of "choosing the sport facilities in a new park" increases when considering some constraints provided by decision makers and the preferences given by a community where some preferences are more representative than others.

In this example, the participants involved in the decision-making process are: a *large number of persons*, who are members of a community; and *multiple decision makers*, namely the city's Mayor and the Councilors.

As could be noticed in these introductory examples, the complexity of a decision is related to the number of persons that are involved, as well as to the diversity of their preferences which might not be sharply defined. These preferences might be expressed using different domains and are based on their knowledge, experience or area of expertise. Here, the challenge is to handle a large number of heterogeneous and imprecise preferences because all the participants are considered to be an important source of information to make better motivated decisions.

Addressing this challenge constitutes the main motivation in this dissertation because these days decision makers seem to be increasingly interested in the opinions (or preferences) given by persons around a community (and sometimes around the world) through different sources including social media channels. This motivation establishes the purpose of the study described in the next section.

1.2 Purpose of the Study

The purpose of this dissertation is to provide a set of tools that helps a decision maker to make better motivated decisions by a proper *handling of a large number of (fuzzy) preferences*, *identifying and evaluating relevant preferences* and *handling multiple perspectives.* Herein, by *'preference'* is meant a greater interest expressed by an individual for a particular alternative over others which might not be sharply defined; by *relevant* is meant a variety of preferences which are significant (or important) to a particular person acting as a decision maker; and by *'perspective'* is understood a position (e.g., social, technical, financial or environmental) adopted by a decision maker when expressing his/her preferences or constraints.

Handling a large number of preferences. Preferences related to a particular decision involving a large number of participants —such as the members of a community— could be gathered through different sources like fan pages, surveys, polls and social network applications. In this case, a decision maker could be overwhelmed by the potential high number of preferences, and hence taking a final decision could become a complex task. Therefore to cope with this challenge, this study aims to provide some sort of simplification mechanism that manages a large number of preferences. For illustration purposes, Figure 1.1 depicts a decision-making process where a large number of participants provide their preferences and a decision maker provides his/her perspective when expressing his/her constraints (or preferences).

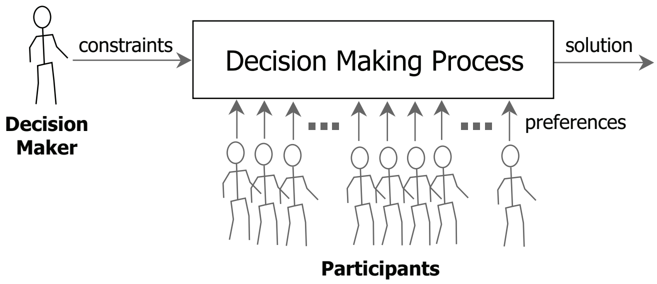


Figure 1.1: A decision-making process that involves a decision maker and a large number of participants.

Identifying and evaluating relevant preferences. In the presence of heterogeneous preferences, these can be categorized from a decision maker's perspective. These preferences are provided by participants having different education levels, areas of expertise and personal profiles while the perspective of a decision maker allows for reflecting a combination of characteristics that make some preferences to be considered relevant or more important than others. In this case, the characteristics that make some preferences more relevant need to be identified and further used for evaluation purposes. As an illustration, Figure 1.2 depicts a decision-making process where a large number of participants provide their preferences, and some participants are considered to be relevant from a decision maker's perspective. Here, the participants considered to be relevant to some extent are depicted with a gray head —the more gray the head is the more relevant the participant is.

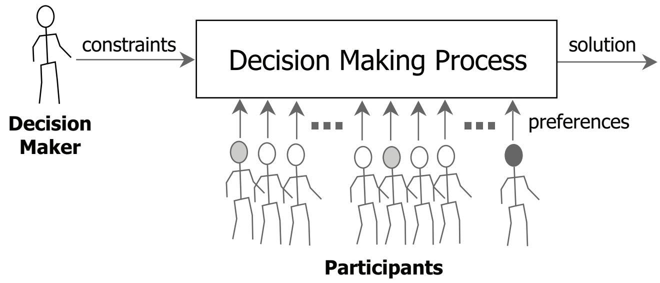


Figure 1.2: A decision-making process that involves a large number of participants where some of them are considered to be relevant (to some extent) according to a decision maker's perspective.

Handling multiple perspectives. Several decision makers could be involved in a decision, acting (most of the time) from a specific perspective (e.g., social, technical, financial or environmental). In this case, it might occur that some relevant information could be ignored from a specific perspective because the information is filtered according to each of the decision maker's preferences or constraints. For instance, a decision maker may prefer the opinions provided by a majority group while other decision maker may prefer the opinions provided by some specific professionals —i.e., some participants are considered being relevant according to a decision maker. Hence, to address this challenge this dissertation aims to provide a tool that merges preferences reflecting multiple perspectives while diminishes information loss.

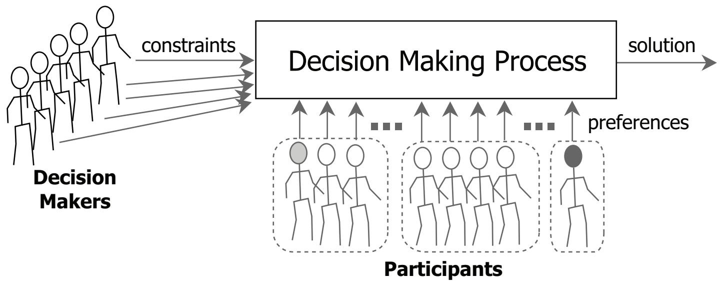


Figure 1.3: A decision-making process that involves several decision makers and a large number of participants. Here, while a decision maker may prefer the opinions of a majority group, another may prefer the opinions provided by relevant participants even though they correspond to a minority group.

1.3 Related Work

Decision making is a process where several persons can participate to make a decision together. In group decision-making (GDM) problems a set of two or more experts $E = \{e_1, e_2, \ldots, e_n\}$ are involved providing their preferences over a set of alternatives $X = \{x_1, \ldots, x_k\}$, where these alternatives are characterized by multiple (and sometimes conflicting) criteria $C = \{c_1, \ldots, c_t\}$. Here, a solution consists in finding an alternative (or a subset of alternatives) that is considered to be acceptable by the group of experts (regarding the criteria).

To solve decision-making problems, the experts can express their preferences using different preference structures —i.e., utility values [1], preference ordering [2] and preference relations [3]. In this context, the most frequently used preference structure is based on preference relations, whereby each expert (participating in a decision) expresses his/her preferences using pairwise comparisons of alternatives. Once all the preferences are available, the group decision-making problem consists in deriving a solution.

A solution can be derived either by the individual preference relations (direct approach [4]) or by computing a social fuzzy preference relation (indirect approach [5, 6]) representing the preferences of the group. Independently of the selected approach, a group decision-making process has two phases [7]: (i) *aggregation* that combines the individual expert preferences; and, (ii) *exploitation* that selects the alternative (or subset of alternatives) for the decision problem.

To solve GDM problems there are different models that can be used and two well known group decision-making models are described as follows.

- *A soft majority approach.* This model considers that "a solution should reflect what a majority of individuals prefer" [8], where a majority might be given through linguistic quantifiers such as 'most' instead of a threshold number of persons. This model can also be referred as *soft consensus* and it uses a collection of individual preference relations where it is possible to derive a solution using either the direct or the indirect approach. New definitions of *degrees of consensus* have been presented in [9, 10, 11] and these are frequently used within a group of experts under fuzzy preferences and fuzzy majorities.

  In [12] is presented soft computing for supporting reaching a consensus. This work is of particular interest considering that an alternative is considered to have a hierarchical structure, based on components and attribute values. For instance, when selecting a car for a taxi corporation, each alternative corresponds to a car model which may be characterized as a whole by different attributes such as engine, color, weight, etc. This allows for different levels of consensus where it is possible that there is no consensus as to the choice of an alternative (e.g., a particular car model), but there is consensus at a different level (e.g., that French cars are preferred). Additionally, this approach uses bipolar preferences provided by the experts, i.e., where the individuals express their level of preference of an alternative over others, as well as the level to which an alternative is not preferred over others.

- *Consensual processes.* The aim of this model is to obtain a solution with a high satisfaction among multiple decision makers (or experts) where the initial preferences might change under the guidance of a moderator in order to approximating it to a collective decision [13, 14, 15]. Although most of the consensual processes consider that a small number of decision makers are involved (i.e., ten decision makers or less), research on large groups of decision makers participating in a consensus process exist and are briefly described next.

  A graphical monitoring tool called MENTOR is presented in [16]. This tool is oriented to give support to the moderator of the consensual process by facilitating the analysis of information regarding the consensus level among a large number of decision makers. The decision-maker's opinions are expressed by means of numerical preference relations and these are visualized as data points in a 2-D space. Another consensus model, presented in [17], classifies large groups of decision makers according to their cooperative and non-cooperative behaviors. This model is based on the decision maker's preference relations where the importance weights of experts with non-cooperative behaviors are penalized (i.e., their weights are reduced during the process). It is worth to mention the linguistic consensus model presented in [18], where the experts' preferences are expressed by means of fuzzy linguistic preference relations, due to its applicability in Web 2.0 communities. This model is based on a delegation scheme in which experts may choose to delegate into other experts (typically) with similar opinions. In this case the similarity (between the opinions) is obtained by computing the distance between the expert's preferences (see Section 3.2.1.2). The delegation scheme allows to manage possible intermittent participation rates (e.g., users that might not collaborate in a consensus round).

The aforementioned-models handle fuzzy preferences provided by the decision-makers (or experts) that are involved in a decision problem. Although it is possible handling imprecise preferences, usually expressed over an alternative or regarding a specific criterion value, it is also desire to handle these preferences over a continuous set of values regarding a criterion. Among these models there are consensual processes that handle a large group of decision makers, however to the best of the author's knowledge these do not handle a large number of regular participants —i.e., persons that might provide their preferences without providing a final decision.

Although there are approaches that focus on a large number of participants such as democratic voting, these are mostly characterized by voting systems where the participants select the preferred alternative. So, preferences are expressed in a discrete way —e.g., an alternative that is or is not preferred is expressed in a binary way. For instance, one may consider an online scheduling system such as Doodle which allows a large number of participants to express their preferences regarding a date/time in order to decide on a mutually agreeable time for an event [19]. However, in the presence of an imprecise concept it is desired the selection of an alternative that is consistent with the participant preferences. For example, one may consider the “level of entertainment” of a film to be imprecise —here, the transition of being ‘entertained’ or not can be gradual.

Other approaches include the *direct and participatory e-democracy* presented in [20] which looks for transparency in the participation process itself on all political levels. This approach considers that “a small part of all citizens have the right to bring any law decided in the parliament (People Referendum) or any proposition for a constitutional or legal change to a referendum (peoples’ vote) to a decision of all people” [20]. While the participatory e-democracy approach takes into account the opinions (or preferences) given by regular participants, the *E-consultation* approach has no direct influence in the decision of a policy because the consulted persons usually are not involved in the decision-making process [21].

1.4 Research Questions

This section provides the following main questions coupled with details regarding their context.

1.4.1 Handling a Large Number of Preferences

As mentioned in the Introduction, the availability of a large number of preferences provided by a community increases the complexity of decision making. Therefore, the following question tries to address the challenge of handling a large number of preferences:

(1) how to simplify the complexity of a decision-making problem that involves a large number of persons in the decision?

It should be noted that someone who is ‘involved’ can express his/her preferences, but he/she might or might not decide on the result. For instance, potential customers of a product, citizens of a country, or students in an educational institution may express their preferences without providing a decision. Furthermore, the involved persons (or participants) might have different levels of knowledge (students, non-experts and professionals), various areas of expertise (engineering, medicine, journalism, among others) and different personal profiles (single, married, parents, among others). Since within a large group might exist different preferences, a specific question to solve is *how to handle heterogeneous preferences?*

1.4.2 Identifying and Evaluating Relevant Opinions

Considering the presence of heterogeneous preferences, a decision maker might be more interested in relevant preferences. Here, relevant preferences are preferences considered to be worthy of extra attention due to they have one or more characteristics that are important for the decision maker. For instance, a decision maker may put more attention to the opinions provided by VIPs[1]. In this case, the decision maker's perspective has an important role since what is deemed to be relevant for a decision maker might be not relevant for another. In this context, the main research questions is

(2) how to identify and evaluate preferences considered being relevant from a decision-makers perspective?

This research question has two specific questions: (i) *how to identify relevant preferences within a large group where some preferences are more representative than others?* and (ii) *how to assist a decision maker to evaluate the preferences that best suits his/her perspective?*

1.4.3 Handling Multiple Perspectives Provided by Decision Makers

As illustrated in the introductory examples, the presence of constraints provided by several decision makers increases even more the complexity of a decision. In this regard, when each decision maker expresses his/her position (or perspective) the following question raises:

(3) How to simplify the complexity of a decision-making problem that includes multiple perspectives in an organizational environment, as well as the preferences given by a large number of persons?

Here, an organizational environment is taken into account considering that, within this environment, several decision makers can go through a decision-making process. In this case, several decision makers might express their particular perspectives (e.g., economical, environmental, financial, etc.) supported by the information that he/she has access to (e.g., surveys, polls, focus groups, etc.). Therefore, two specific questions to address are the following: (i) *how to split a complex multiple-perspective decision-making problem into several simple single-perspective ones?* and (ii) *how to combine preferences from multiple perspectives?*

1.5 Scope

Decision making is a process that relates to different study areas, but within this dissertation it relates to computational intelligence, and more specifically the one that relies on soft computing. By *soft computing* is meant a set of techniques that "mimic the ability of the human mind to effectively employ modes of reasoning that are approximate rather than exact" [22], which allows for obtaining low cost solutions when imprecision and subjectivity are present.

In this dissertation, soft computing techniques facilitate the preference modeling when the preferences are imprecisely stated by the persons involved in a

[1] *VIP* is an acronym that stands for very important person.

decision. Although in the presence of a large number of preferences these might be provided using different domains (i.e., numerical, interval-valued, linguistic), within this dissertation it is considered that all the preferences could be unified in a common domain as it will be further explained. In this way, this dissertation focuses primarily on the understanding of a novel approach to a decision-making problem where a community becomes an important source of information for decision makers acting from multiple perspectives (e.g., social, technical, financial or environmental). Moreover, considering that decision makers use information arriving from different sources, this dissertation proposes an information fusion technique as a tool to help the decision makers to make better motivated decisions.

1.6 Significance of the Study

The significance of the study is demonstrated by briefly describing some of the existing decision-making models and how these assist a decision maker managing the complexity in a decision. For illustration purposes the following example is used.

A decision about a new ice cream. An ice cream company has to decide on a new ice cream (product), based on its flavor and its corresponding nutritional information (criteria). This company is represented by the General Manager who can make a final decision taking into account the preferences given by two intermediate managers, i.e. the Product and Marketing Managers, with respect to their competence area or perspective. The perspective of the Product Manager is mainly focused on (using) the available facilities of the company and (how might be perceived) the nutritional information of the new ice cream, while the perspective of the Marketing Manager is focused on the flavor of the new sweet. The managers of this company would like to consider the preferences given by potential customers of the product through the company's social media channels (e.g., twitter, instagram, facebook).

In the given example, an option corresponds to a specific ice cream flavor with its nutritional information.

In the case of using a *soft majority approach*, the product might be selected by looking for the best acceptable option by *most* of the managers, i.e. what a majority of the managers prefer. Analogously, the intermediate managers may select the option that a majority of potential customers suggest. Here, it is possible to collect pairwise comparisons among the options given by the managers, but those given by potential customers might not be available by means of social media. Social media users usually express their preferences through small posts or hashtags[2]. Hence a further step for handling a large number of preferences given by persons that are involved (and not necessarily decide on the result) should be further studied.

[2] "A word or phrase preceded by a hash sign (#), used on social media sites such as Twitter to identify messages on a specific topic" [23]. For example, a twitter user may post "Craving for a #chocoberry #icecream!", "#bananauts love!", or "#votebananaberry ".

In the case of using a *consensual process*, managers can participate in a process looking for a consensual decision, but it is more complex when including a large number of potential customers to participate —especially considering that a unique flavor may not satisfy all the potential customers. Here, one can bear in mind that the managers might explore other possibilities (from different perspectives) like a new target market based on a variety of preferences—or additional information like age, gender or location.

As could be noticed, handling a large number of (fuzzy) preferences given by persons involved in a decision is a more complex problem that needs research attention especially these days when different technologies facilitate the access to a variety of information sources (e.g., fan pages, surveys, polls, social media applications, among others) including citizens as an important source of information by providing a diversity of preferences. This aspect can be seen as an opportunity to handle the preference variety within big data problems.

Besides a large number of preferences which might not be precisely defined, a decision may become more complex when multiple perspectives coexist. In this way, what is considered to be relevant for a decision maker sometimes might be considered as irrelevant (or partially relevant) for other decision makers. For example, some decision makers may disregard a minority group with an "unusual" preference, while other managers in the search of a new target group could pay more attention to this group. Additionally, multiple perspectives provided by peers within organizations that operate in different places (e.g., in more than one country) may be reached. Therefore, cases where the headquarters would like to take into account the opinions given by their organizational units (e.g., regional and sub-regional units) and their respective customers could be also further studied.

To put it briefly, one can consider this dissertation worthy of attention because more complex decision-making problems that include fuzzy preferences given by a community (i.e., a large number of persons) are considered important, and most of the time several decision makers are involved providing multiple perspectives (e.g., social, technical, financial or environmental) to make better motivated decisions —considering that seldom a single decision maker has knowledge of the entire domain of a decision problem [24]. Therefore the need to provide a set of tools that helps decision makers to make better supported decisions by a proper handling of a large number of (fuzzy) preferences and multiple perspectives is foreseen. Moreover, different organizations (e.g., charitable, governmental, non-profit or business organizations) may enhance the opportunities of application areas in a decision-making context including environmental issues [25], e-democracy [26], suitability maps [27] and water resources management [28].

1.7 Dissertation Outline

This dissertation has 6 chapters and most of them are based on articles that have been published either in international peer reviewed journals, international peer reviewed conference proceedings, or as chapters in books revised and edited by leading experts in the computational intelligence field. For the sake of referencing, the corresponding articles are explicitly mentioned at the beginning of each chapter.

Next, the chapters of this dissertation are briefly described:

Chapter 2 covers preliminary concepts that facilitate the understanding of the following chapters. These include concepts on fuzzy sets and how these may assist a person to define criteria in the area of decision making. Additionally, this chapter presents concepts on linguistic computation and aggregation operators. Although this chapter corresponds to preliminary concepts required to properly understand the remaining chapters, experts in the area of fuzzy theory and decision-making may skip some parts of this chapter. Thus, this chapter may be revised in whole or in part for the purpose of being used herein as a reference tool. In particular, this chapter constitutes a brief collection of contributions made by different authors, however the examples and further explanations are contributions of the author of this dissertation.

Chapter 3 presents a method that aims to reduce the complexity of a decision-making problem when it involves a large number of persons expressing their opinions (or preferences). For instance, potential customers of a product, citizens of a country, or students in an educational institution. Bearing in mind that an strategy to solve a decision-making problem is that each person (e.g., a citizen, an expert or a decision maker) expresses his/her preference over a specific criterion as a matter of degree (i.e., by means of a membership function), consequently the total number of membership functions equals to the number of persons involved. However, if there are some persons with similar opinions, it is possible to reduce the amount of preferences (and hence the complexity) by considering the similarity among their corresponding membership functions. As a novelty, the similarity approach presented in this chapter is a method that uses a symbolic notation to depict each membership function taking into account its shape characteristics –e.g., slopes, preference levels, core segments, left and right spreads.

Chapter 4 proposes a methodology to determine relevant opinions (or preferences) according to a decision maker's perspective. This chapter focuses on the identification and evaluation of relevant preferences based on some characteristics over groups of similar preferences. The characteristics that are considered within this chapter include a *cohesion measure* and the *representativeness* of the group, where the cohesion of a group is a measure that takes into account the level of togetherness among its contained membership functions; and the representativeness combines the number of membership functions and the number of noticeable represented preferences (i.e., preferences in the group that are worthy some extra attention or are more important).

Section 5 presents a *novel decision-making model* that allows for combining preferences given by persons having different perspectives (e.g., economical, technical, environmental, etc.), including decision makers, and aimed to be suitable for different organizational structures (e.g., multilevel structures). Furthermore, bearing in mind the aim of a flexible model, this chapter introduces the *decision-making unit (DMU)* concept as a primary component that facilitates the propagation of preferences throughout an organizational structure in order to make better motivated decisions.

Chapter 6 outlines the main contributions of the research leading up to this dissertation and settles down some opportunities for further work. Then, the Conclusions of this dissertation are presented, followed by some open issues that have been raised.

References

[1] Tetsuzo Tanino. *On group decision making under fuzzy preferences.* In Multiperson Decision Making Models using Fuzzy Sets and Possibility Theory, pages 172–185. Springer, 1990.

[2] Shuwei Chen, Jun Liu, Hui Wang, and Juan Carlos Augusto. *Ordering based decision making–a survey.* Information Fusion, 14(4):521–531, 2013.

[3] S.A. Orlovsky. *Decision-making with a fuzzy preference relation.* Fuzzy Sets and Systems, 1(3):155–167, 1978.

[4] Francisco Herrera, Enrique Herrera-Viedma, and José Luis Verdegay. *Direct approach processes in group decision making using linguistic OWA operators.* Fuzzy Sets and systems, 79(2):175–190, 1996.

[5] Francisco Chiclana, Francisco Herrera, and Enrique Herrera-Viedma. *Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations.* Fuzzy sets and Systems, 97(1):33–48, 1998.

[6] Enrique Herrera-Viedma, Francisco Herrera, and Francisco Chiclana. *A Consensus Model for Multiperson Decision Making With Different Preference Structures.* 32(3):394–402, 2002.

[7] Marc Roubens. *Fuzzy sets and decision analysis.* Fuzzy Sets and Systems, 90(2):199–206, 1997.

[8] Janusz Kacprzyk. *Group decision making with a fuzzy linguistic majority.* Fuzzy Sets and Systems, 18(2):105–118, mar 1986.

[9] Janusz Kacprzyk and Mario Fedrizzi. *A 'soft' measure of consensus in the setting of partial (fuzzy) preferences.* European Journal of Operational Research, 34(3):316–325, 1988.

[10] Janusz Kacprzyk and Mario Fedrizzi. *A 'human-consistent' degree of consensus based on fuzzy login with linguistic quantifiers.* Mathematical Social Sciences, 18(3):275–290, 1989.

[11] Mario Fedrizzi, Janusz Kacprzyk, and Hannu Nurmi. *Consensus degrees under fuzzy majorities and fuzzy preferences using OWA (ordered weighted average) operators.* Control and Cybernetics, 22(4):71–80, 1993.

[12] Janusz Kacprzyk and Sławomir Zadrożny. *Soft computing and web intelligence for supporting consensus reaching.* Soft Computing, 14(8):833–846, 2010.

[13] Enrique Herrera-Viedma, José Luis García-Lapresta, Janusz Kacprzyk, Mario Fedrizzi, Hannu Nurmi, and Sławomir Zadrożny. *Consensual processes*, volume 267. Springer, 2011.

[14] Steven Saint and James R. Lawson. *Rules for reaching consensus: a modern approach to decision making.* Pfeiffer, 1994.

[15] Lawrence E Susskind, Sarah McKearnen, and Jennifer Thomas-Lamar. *The consensus building handbook: A comprehensive guide to reaching agreement.* Sage Publications, 1999.

[16] Iván Palomares, Luis Martínez, and Francisco Herrera. *MENTOR: A graphical monitoring tool of preferences evolution in large-scale group decision making.* Knowledge-Based Systems, 58:66–74, 2014.

[17] Iván Palomares, Luis Martínez, and Francisco Herrera. *A consensus model to detect and manage non-cooperative behaviors in large scale group decision making.* IEEE Transactions on Fuzzy Systems, 22(3):516–530, 2014.

[18] Sergio Alonso, Ignacio J. Pérez, Francisco Javier Cabrerizo, and Enrique Herrera-Viedma. *A linguistic consensus model for Web 2.0 communities.* Applied Soft Computing, 13(1):149–157, January 2013.

[19] Katharina Reinecke, Minh Khoa Nguyen, Abraham Bernstein, Michael Näf, and Krzysztof Z. Gajos. *Doodle around the world: online scheduling behavior reflects cultural differences in time perception and group decision-making.* In Proceedings of the 2013 Conference on Computer Supported Cooperative Work, pages 45–54. ACM, 2013.

[20] Amr Huber. *E-Democracy in a participatory form of democracy (direct democracy).* Sofia CAHDE. Retrieved July, 7:2009, 2007.

[21] Klaus Petrik. *Participation and e-democracy how to utilize web 2.0 for policy decision-making.* In Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government, pages 254–263. Digital Government Society of North America, 2009.

[22] Lotfi Zadeh. *Fuzzy logic, neural networks, and soft computing.* Communications of the ACM, 37(3):77–84, 1994.

[23] Oxford Dictionaries. *"Hashtag".* n.d. Web. 6 February 2016. Retrieved from http://www.oxforddictionaries.com/definition/english/hashtag.

[24] Witold Pedrycz, Petr Ekel, and Roberta Parreiras. *Fuzzy multicriteria decision-making: models, methods and applications.* Wiley, United Kingdom, 1st edition, 2011.

[25] Maria Franca Norese. *ELECTRE III as a support for participatory decision-making on the localisation of waste-treatment plants.* Land Use Policy, 23(1):76–85, 2006.

[26] Jinbaek Kim. *A model and case for supporting participatory public decision making in e-democracy.* Group Decision and Negotiation, 17(3):179–193, 2008.

[27] Guy De Tré, Jozo Dujmović, and Nico Van De Weghe. *Supporting spatial decision making by means of suitability maps.* Uncertainty Approaches for Spatial Data Modeling and Processing, 281:9–27, 2010.

[28] Piero Fraternali, Andrea Castelletti, Rodolfo Soncini-Sessa, Carmen Vaca, and Andrea Emilio Rizzoli. *Putting humans in the loop: Social computing for Water Resources Management.* Environmental Modelling & Software, 37:68–77, 2012.

Chapter 2

Preliminaries

This chapter constitutes a brief collection of basic concepts that facilitate the understanding of the remaining chapters, and might be used herein as a reference tool. Although this chapter constitutes a brief collection of contributions made by different authors, the examples and further explanations are contributions of the author of this dissertation.

2.1 Introduction

Decision making is a process where imprecision and subjectivity are present, especially when persons involved in a decision problem express their preferences. Under these circumstances, soft computing allows for obtaining low cost solutions based on *fuzzy logic*.

Fuzzy logic [1] is a form of logic that enable elements to be part of a *fuzzy set* to a certain grade. For example, the "usefulness" of a product might be considered as a subjective feature, and thus a grade of membership (between 0 and 1) can be given by potential customers —where 0 denotes no usefulness, 1 denotes complete usefulness and different values between these denote a partial level of usefulness. Therefore, Section 2.2 presents basic concepts on fuzzy sets for a better understanding of their role in representing preferences given by individuals.

Preferences given by persons with different personal profiles (e.g., singles, married, parents), different levels of knowledge (e.g., non-experts, students and professionals) and a variety of expertise areas (e.g., engineering, medicine, journalism, among others) might be provided using different domains. For instance, a person may use the following expressions to denote his/her preferences over the criterion *weight* of a product like a *dumbbell*: "I prefer a product of 2 kilograms" (numerical value), "I prefer a product between 1.5 and 2 kilograms" (interval-valued), or "I prefer a *light* product" (linguistic approach). However preferences expressed using different domains (i.e., numerical, interval-valued, linguistic) could be unified by means of fuzzy sets in a common domain[1]. For

[1] The interested reader may refer to [2] for details on different transformation functions

this reason, Section 2.3 describes the use of fuzzy sets in the definition of criteria. This includes modeling preferences from a criterion value —given by participants of a decision process—, defining the attribute criterion in a piecewise linear function such as a triangular or a trapezoidal membership function.

Among trapezoidal membership functions it is feasible to differentiate some characteristic shapes (when representing preferences) which could be used to facilitate function comparisons from a human's perspective. Hence Section 2.4 presents concepts on linguistic computing as an approximate technique to represent qualitative aspects using words.

In Section 2.5 the concept of an aggregation function, or aggregation operator in the fuzzy set context, is presented to further understand their use when reflecting the perspective (on criteria) of a decision maker in a desirable way. Bearing in mind that the availability of aggregation functions is broad[2], herein only some well-known classes of aggregation functions are described.

2.2 Basic Concepts on Fuzzy Sets

A *fuzzy set* is a concept, fruitfully generalized, from the basic mathematical concept of a set [5]. In classical set theory an element belongs or not to a set, while in fuzzy set theory the "belongingness" to a set is a matter of degree [6]. In this way, a fuzzy set may represent a category (or a concept) where a transition from *full* to *none belongingness* is perceived, and therefore it is possible that "elements of a universe can be members of a class and at the same time belong to other classes with different degrees" [6]. For example, when considering concepts with regard to temperature there are imprecise boundaries between what is considered *hot*, *warm* and *cold*, and a specific temperature might be considered to some extent *hot* and at the same time *warm* to some extent. In this example, fuzzy sets facilitate their representation and interpretation in a given context like "room temperature".

According to Zadeh, a fuzzy set A is a class of objects with a continuum of grades of membership characterized by a membership function μ_A [5], where a membership function is formally defined as follows:

Definition 2.1 (Membership Function [5])
A membership function μ_A over a universe of discourse X is a mapping $\mu_A : X \mapsto [0,1]$ that associates each $x \in X$ with a real number $\mu_A(x)$ in the unit interval $[0,1]$ to represent the grade of membership of x in A. Values that are closer to 1 denote higher grades of membership, while values that are closer to 0 denote lower grades of membership.

and operators to unify information from different domains (i.e., numerical, interval-valued, linguistic), and may also refer to a complementary survey on the fusion process with heterogeneous preference structures in group decision-making [3].

[2] The interested reader may refer to [4] for a deep analysis of aggregation functions, their corresponding properties and their interpretation.

Definition 2.1 allows that any function $\mu_A : X \mapsto [0,1]$ may correspond to a membership function representing a fuzzy set A, and consequently graphical representations of membership functions come in different shapes including triangular, trapezoidal, Gaussian and S-membership functions. However, a membership function "should reflect the perception (semantics) of the concept to be represented and further used in problem solving" [6], and therefore should be selected with caution.

To select a membership function, representing a fuzzy set, there are certain characteristics that can be considered [7], among them the following:

- The *support* of a fuzzy set A, denoted by $supp(A)$, over a universe of discourse X is a set that contains all the elements in X with nonzero membership grades in A. That is,

$$supp(A) = \{x \in X |\ \mu_A(x) > 0\}.$$

- The *core* of a fuzzy set A, denoted by $core(A)$, over a universe of discourse X is a set that contains all the elements in X with membership grades equal to one. That is,

$$core(A) = \{x \in X |\ \mu_A(x) = 1\}.$$

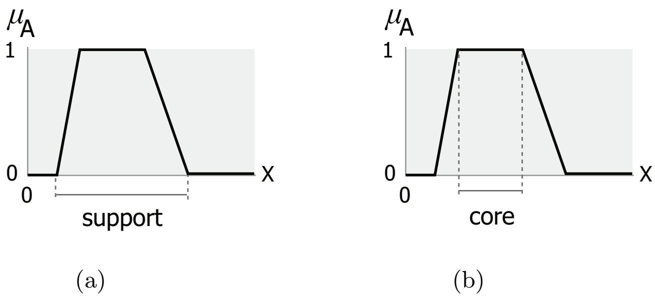


Figure 2.1: Fuzzy set A represented by membership function μ_A and its characteristics. (a) The *support* denoted by $supp(A)$, and (b) the *core* denoted by $core(A)$.

- The *height* of a fuzzy set A, denoted by $hgt(A)$, is the largest membership grade obtained by any element in the set. In the literature, the height can be also be referred as the supremum, *sup*, as shown in the expression

$$hgt(A) = \sup_{x \in X} A(x).$$

- The *normality* of a fuzzy set is given by its height. Thus, a fuzzy set A is called *normal* if $hgt(A) = 1$, and it is called *subnormal* if its height is lower than one —i.e., $hgt(A) < 1$.

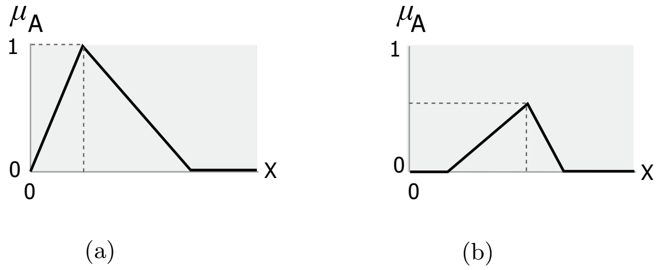


Figure 2.2: (a) A *normal* fuzzy set A. (b) A *subnormal* fuzzy set A.

In the remaining of this dissertation normalized fuzzy sets are used, therefore a normalized membership function is formally defined as follows.

Definition 2.2 (Normalized Membership Function)
A normalized membership function μ is a membership function over a universe of discourse X such that $\mu(x) = 1$ holds for at least one $x \in X$.

Additionally, herein trapezoidal and triangular membership functions have been selected to represent preferences on the criteria given by individuals (or experts). This selection has two main advantages: (i) these membership functions could be built with only a few parameters, and (ii) they are widely known and frequently used for representing linguistic terms [7]. So, these membership functions are more detailed next.

2.2.1 Trapezoidal Membership Functions

These are piecewise linear functions defined in the universe of real numbers that could be built through parameters a, b, c, and d as follows:

$$\mu_A(x) = \begin{cases} 0 & , \quad x \leq a \\ \frac{x-a}{b-a} & , \quad a < x < b \\ 1 & , \quad b \leq x \leq c \\ \frac{d-x}{d-c} & , \quad c < x < d \\ 0 & , \quad x \geq d \end{cases} \tag{2.1}$$

In Equation 2.1, parameters a, b, c, and d satisfy $a \leq b \leq c \leq d$, and graphically represent dividing points among the segments of a trapezium as shown in Figure 2.3.

It is worth to mention that depending on the values of parameters a, b, c and d, different trapezium shapes can be obtained. For instance, pairs of parameters with equal values such as $c = d$, $a = b$ and $b = c$ are illustrated in Figure 2.4. Hereafter, for readability purposes, a trapezoidal membership function μ_A representing a fuzzy set A may also be denoted as a 4-tuple (a, b, c, d) where these parameters correspond to the ones specified in Equation 2.1.

It could be noticed that trapezoidal membership functions where b and c have equal values correspond to functions represented by a triangular shape.

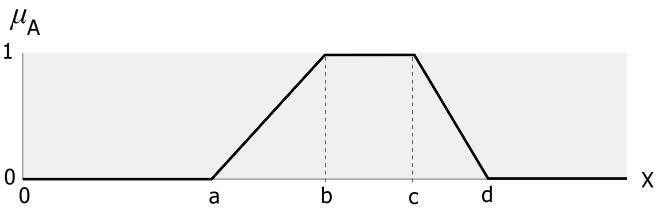


Figure 2.3: Trapezoidal membership function μ_A denoted by parameters *a, b, c* and *d*.

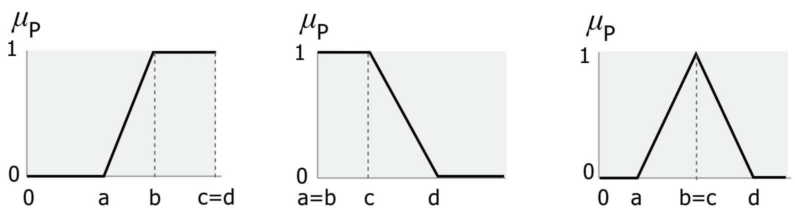


Figure 2.4: Characteristic shapes in trapezoidal membership functions where pairs of parameters have equal values, among them trapeziums where $c = d$, $a = b$ and $b = c$.

2.2.2 Triangular Membership Functions

These are piecewise linear functions defined in the universe of real numbers that could also be built using Equation 2.1 with the condition that b and c have equal values. In a triangular membership function, the a and d values correspond to the lower and upper bounds respectively, while the $b = c$ value denotes a *typical value*.

With the purpose of illustration, one can consider the following definition of *room temperature* given by the Oxford English Dictionary: “A comfortable ambient temperature, generally taken as about 20°C”[3].

Based on the aforementioned definition, the concept of *room temperature* can be perceived as imprecise and therefore a fuzzy set may be used to facilitate its representation. Considering that the given 20°C corresponds to the typical value, one can consider the use of a triangular membership function where the spread given by the lower and upper bounds (given by parameters a and d respectively) may change from one individual to other.

For instance, the fuzzy set *temp* represented by the membership function $\mu_{\text{temp}}(x)$ is shown in Figure 2.5, where $a = 15$°C, $b = c = 20$°C and $d = 25$°C.

[3] “Room temperature”. Oxford Dictionaries. Oxford University Press, n.d. Web. 26 January 2016. Retrieved from http://www.oxforddictionaries.com/definition/english/room-temperature.

This representation denotes that temperatures below 15°C and those above 25°C do not correspond to the comfortable room temperature concept, while 20°C completely denotes the concept of comfortable room temperature.

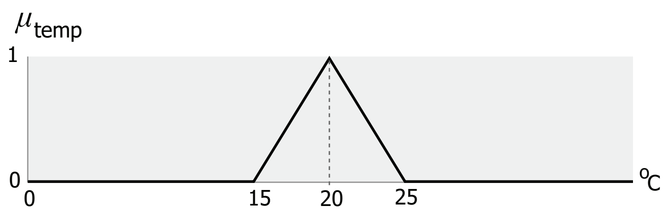


Figure 2.5: Triangular membership function μ_{temp}, denoted by the 4-tuple $(15, 20, 20, 25)$, representing the concept of *comfortable room temperature*.

2.3 Definition of Criteria

In a decision-making context, criteria may be defined by means of fuzzy sets [8] where membership grades express levels of preference on the criteria [9]. Accordingly, when membership reflects preference [10], a membership function $\mu_A(x)$ represents a set of more or less preferred values of a decision variable x of a universe X where $\mu_A(x)$ represents the preference level in favor of value x. Therefore, fuzzy sets may represent the preferences given by individuals (i.e., experts, decision makers or members of a community) over criteria, where the levels of preference may differ from one individual to another. However, these levels of preference can be transformed to a common scale such as the unit interval. Hence, hereafter in this dissertation it is assumed the use of normalized membership functions (Definition 2.2).

Using computational intelligence techniques, a person could express his/her preferences using a fuzzy set P with respect to a specific criterion over the values of a universe X as a matter of degree. That is, $0 \leq \mu_P(x) \leq 1$ where 0 denotes the lowest preference level, 1 denotes the highest level of preference, and different values in between denote a partial level of preference on the value x. For instance, a person may express his/her preferred temperature for a room using the fuzzy set *temp* depicted in Figure 2.5. In this case, the highest level of preference $\mu_{\text{temp}}(x) = 1$ is given when $x = 20$°C.

For the purpose of illustration, Figure 2.6 shows a triangular membership function μ_P, given by the 4-tuple (a, b, c, d), where the lowest preference level $\mu_P(x) = 0$ is given when $x < a$ or $x > d$, the highest level of preference $\mu_P(x) = 1$ is given when $x = b = c$, and other intermediate preference levels in $]0,1[$ are given when $b > x > a$ or $d > x > b$.

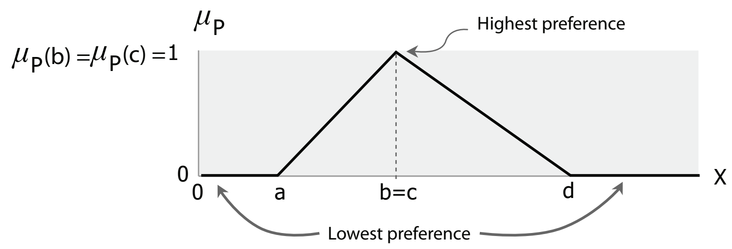


Figure 2.6: Triangular membership function denoting preference levels $\mu_P(x)$ on the x values.

2.3.1 Preference Modeling from Criterion Values

Some decision-making problems aim to involve a large number of participants that not necessarily act as decision makers and only provide information regarding their preferences on the criterion values under consideration. In these cases, the preferences given by a person can be modeled through a fuzzy set P as long as he/she provides some values that will be used for defining the attribute criterion in a membership function μ_P.

2.3.1.1 Specifying an *Ideal Value*

A person that expresses his/her preference on a given criterion by specifying an "ideal value" can represent this preference by means of a triangular membership function. Here, the highest level of preference corresponds to the point where x equals the *ideal value*, and the spread between $x = a$ and $x = d$ may vary to denote the lower and upper bounds respectively. In other words, a triangular membership function depicts the highest preference for a specific value $b = c$ and the lowest preference for values "below a" and values "above d".

For example, a person can express that the temperature "20°C" completely represent the concept of *comfortable room temperature*, while other temperatures "between 15°C and 25°C" partially represent this concept. Moreover, this person can express that values "below 15°C" and "above 25°C" do not represent this concept. Figure 2.5 depicts these preferences by means of membership function μ_{temp}.

2.3.1.2 Specifying a *Range of Values*

A person that expresses the highest level of preference by a range of values can model his/her preferences by means of a trapezoidal membership function. In this case, the highest level of preference is given "between values b and c" and the lowest level of preference may vary.

For example, a person can express that temperatures "between 18°C and 20°C" completely represent the concept of *comfortable room temperature*, while values "below 15°C" and values "above 25°C" do not represent this concept (Figure 2.7).

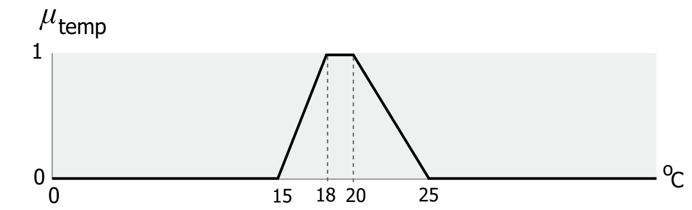


Figure 2.7: A trapezoidal membership function representing the concept of *room temperature* using a range of values "between 18°C and 20°C".

In addition to defining criteria using a numerical or an interval-valued domain, some persons may prefer using a linguistic approach. Therefore, an example using a linguistic approach is presented next, and more details regarding this topic are provided in Section 2.4.

2.3.1.3 Specifying *Linguistic Terms*

Using a linguistic approach, a person can model his/her preferences by means of words. For instance, a person may refer to the concept 'room temperature' with words such as *cold, comfortable* or *warm.* Considering that there are imprecise boundaries between what is considered *cold, comfortable* and *warm*, fuzzy sets can facilitate the representation of these linguistic terms in a given context such as 'room temperature'.

For illustration purposes, Figure 2.8 depicts triangular membership functions $\mu_{\text{cold}}, \mu_{\text{comfortable}}$ and μ_{warm} representing the linguistic terms *cold, comfortable* and *warm* respectively. Here, it could be noticed that a specific temperature such as 17.5°C might be considered to some extent *cold* and *comfortable* at the same time, i.e. $\mu_{\text{cold}}(17.5) = \mu_{\text{comfortable}}(17.5) = 0.5$.

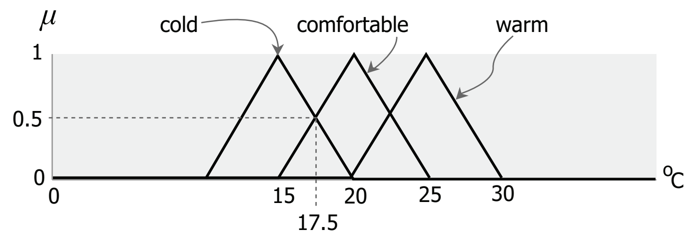


Figure 2.8: Triangular membership functions $\mu_{\text{cold}}, \mu_{\text{comfortable}}$ and μ_{warm} representing the linguistic terms *cold, comfortable* and *warm* respectively.

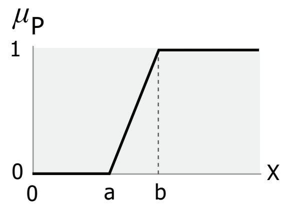
Considering that decision-making problems frequently deal with multiple (and sometimes conflicting) criteria given by decision makers, each criterion can be defined as presented herein in order to model the preferences (or constraints) given by a decision maker.

Next, some characteristic shapes that could be found when modeling preferences are presented considering that the preference levels may vary from one person to other.

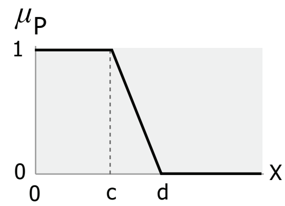
2.3.2 Characteristic Shapes in Membership Functions

In a decision-making context, from the preference point of view, a trapezoidal membership function represents the preference level $\mu_P(x)$ in favor of value x. Consequently, it is feasible to differentiate different characteristic shapes.

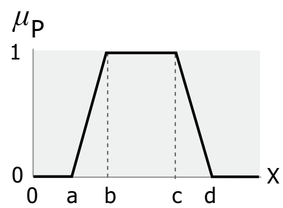
Some of the frequently used characteristic shapes are shown in Figure 2.9 followed by a brief description regarding the preference levels by means of parameters *a,b,c* and *d*.



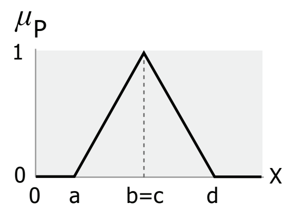
(a) Shape 1.



(b) Shape 2.



(c) Shape 3.



(d) Shape 4.

Figure 2.9: Characteristic shapes in trapezoidal membership functions.

Shape 1, depicts the highest preference for values “above b” and the lowest preference for values “below a” (Figure 2.9a).

Shape 2, depicts the highest preference for values “below c” and the lowest preference for values “above d” (Figure 2.9b).

Shape 3, depicts the highest preference for values “between b and c” and the lowest preference for values “below a” and values “above d” (Figure 2.9c).

Shape 4, depicts the highest preference for a specific value $b = c$ and the lowest preference for values “below a” and values “above d” (Figure 2.9d). It could be noticed that this is a special case of *Shape 3*.

Based on the aforementioned characteristic shapes, a person can use trapezoidal membership functions by specifying the minimum value of $x = a$ where his/her preference starts to increase in order to reach the highest preference level where $x = b$ (Shape 1). Analogously, a person may specify the maximum value of $x = c$ where his/her preference starts to decrease until the point $x = d$ where reaches the lowest preference level (Shape 2). Moreover, a combination of these two cases allows a person to specify the highest level of preference between $x = b$ and $x = c$ values (Shape 3). In the case that the highest preference corresponds to a specific value $b = c$, while the lowest preference correspond to values "below a" and values "above d", it could be noticed that a triangular membership function can be used (Shape 4).

2.4 Linguistic Computation

Linguistic computation allows for representing qualitative aspects using words, in a natural or artificial language, instead of numerical values [11, 12, 13]. So, one can use a *linguistic variable* to characterize different situations where its value is chosen among words. For instance, *cold, comfortable* and *warm* are values referring to the linguistic variable *temperature*. It could be noticed that, in contrast to a numerical variable, linguistic variables allow for the representation of concepts with unclear boundaries.

Formally, the definition of a *linguistic variable* is as follows.

Definition 2.3 (Linguistic variable [11])
A linguistic variable is characterized by a quintuple $< V, T(V), U, G, M >$ *where its components are: (i)* V*, the variable name; (ii)* $T(V)$*, a term set that includes the allowed labels* L *and the linguistic values for* V*; (iii)* U*, the universe of discourse; (iv)* G*, a syntactic rule for generating the names in* $T(V)$*; and (v)* M*, a semantic rule that associates to each linguistic term* X *a meaning* $M(X)$*, where* $M(X)$ *corresponds to a fuzzy subset of* U*.*

For the purpose of illustration, one may consider the linguistic variable *temperature* followed by the identification of the components detailed in Definition 2.3.

- V = temperature, is the variable name;
- $T(V)$ = {very-cold, cold, comfortable, warm, hot}, is the term set;
- $U = [0^{\circ}\mathrm{C}, 40^{\circ}\mathrm{C}]$, is the universe of discourse;
- $M(\text{very-cold}) = \mu_{\text{very-cold}}$, $M(\text{cold}) = \mu_{\text{cold}}$, $M(\text{comfortable}) = \mu_{\text{comfortable}}$, $M(\text{warm}) = \mu_{\text{warm}}$, and $M(\text{hot}) = \mu_{\text{hot}}$.

In the given example, it could be noticed that all the values of a linguistic variable constitute its *term set*, and the *semantic rule* or meaning of a linguistic variable is a fuzzy subset which could be represented by membership functions as shown in Figure 2.10.

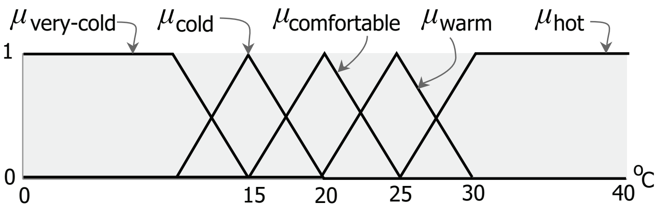


Figure 2.10: Trapezoidal membership functions $\mu_{\text{very-cold}}$, μ_{cold}, $\mu_{\text{comfortable}}$, μ_{warm} and μ_{hot} representing the linguistic terms *very-cold, cold, comfortable, warm* and *hot* respectively.

Although theoretically a term-set could have an infinite number of elements, it is important to pay attention that according to Miller's studies [14] there is an upper limit on our capacity to process information with reliability and accuracy. Thus, when considering a number of terms to denote a concept such as temperature, this limit is seven plus or minus two. Therefore, a linguistic term set should be defined with caution and should not contain too many terms.

2.5 Aggregation

Aggregation is the process of combining several numerical values into a single representative value [4]. Within this dissertation, some aggregation functions are used to combine numerical values in a desirable way to reflect a decision maker's point of view over criteria.

The definition of an aggregation function, taken from [4], has the domain $\mathbb{I}$ which is a nonempty real interval while the integer n represents the number of its variables as follows:

Definition 2.4 (Aggregation Function [4])
An aggregation function *in* $\mathbb{I}$ *is a function* $A^{(n)}(x) : \mathbb{I}^n \to \mathbb{I}$ *that*

(i) is non-decreasing (in each variable)

(ii) fulfills the boundary conditions

$$\inf_{x \in \mathbb{I}^n} A^{(n)}(x) = \inf \mathbb{I} \quad \textit{and} \quad \sup_{x \in \mathbb{I}^n} A^{(n)}(x) = \sup \mathbb{I}.$$

In the fuzzy set context, an *aggregation operator* is used to combine two or more fuzzy sets in a desirable way into one single fuzzy set. An aggregation operator over a finite number of fuzzy sets $n \in \mathbb{N} \backslash \{0\}$, all defined over the same universe of discourse X, can generally be specified by means of a function $h : [0,1]^n \mapsto [0,1]$.

Based on Definition 2.4, there are several aggregation operators that might be used such as the triangular norms and conorms, aggregation operations such as the generalized averages and the ordered weighted averages, and the generalized conjunction/disjunction. In this section these well-known aggregation operators classes are briefly presented.

Although the aforementioned classes of aggregation functions are frequently used, here it is worth to mention others that can be used as well. Among them, the work presented in [15] which provides a study on penalty-based aggregation functions, parametric classes of generalized conjunction and disjunction operations available in the case that either the commutativity or the associativity properties are considered [16].

2.5.1 Triangular Norms and Conorms

The union and intersection operations for sets can be generalized for fuzzy sets for combining two or more fuzzy sets into one single fuzzy set. Generally, the union and intersection operations are specified by means of a binary operation which is defined over the unit interval $[0,1]$ and satisfies some given conditions.

As such, the *intersection* of two fuzzy sets A and B —defined over the same universe of discourse X— can generally be specified by means of a function

$$T : [0,1] \times [0,1] \mapsto [0,1]$$

which takes the membership grades of an element $x \in X$ in the fuzzy sets A and B as arguments and computes the membership grade of x in the intersection of A and B, i.e.

$$\forall x \in X : \mu_{A \cap B} = T(\mu_A, \mu_B).$$

To be intuitively acceptable as an intersection function, the function T must moreover satisfy the following axioms:

- Axiom T1. $\forall a \in [0,1] : T(a,1) = a$ (border condition).
- Axiom T2. $\forall a,b,d \in [0,1] : b \leq d \Rightarrow T(a,b) \leq T(a,d)$ (monotonicity).
- Axiom T3. $\forall a,b \in [0,1] : T(a,b) = T(b,a)$ (commutativity).
- Axiom T4. $\forall a,b,d \in [0,1] : T(a,T(b,d)) = T(T(a,b),d)$ (associativity).

Functions T that satisfy the previous specification and axioms are known in the literature under the name of *triangular norms* or *t-norms* for short.

The *union* of two fuzzy sets A and B —defined over the same universe of discourse X— can also generally be specified by means of a function

$$S : [0,1] \times [0,1] \mapsto [0,1]$$

which takes the membership grades of an element $x \in X$ in the fuzzy sets A and B as arguments and computes the membership grade of x in the union of A and B, i.e.

$$\forall x \in X : \mu_{A \cup B} = S(\mu_A, \mu_B).$$

In order to be intuitively acceptable as a union function, the function S must satisfy the following axioms:

- Axiom S1. $\forall a \in [0,1] : S(a,0) = a$ (border condition).
- Axiom S2. $\forall a,b,d \in [0,1] : b \leq d \Rightarrow S(a,b) \leq S(a,d)$ (monotonicity).
- Axiom S3. $\forall a,b \in [0,1] : S(a,b) = S(b,a)$ (commutativity).
- Axiom S4. $\forall a,b,d \in [0,1] : S(a,S(b,d)) = S(S(a,b),d)$ (associativity).

Functions S that satisfy the previous specification and axioms are known in the literature under the name of *triangular conorms* or *t-conorms* for short.

For t-norms and t-conorms the following inequalities hold:

$$T(a,b) \leq a \leq S(a,b)$$

$$T(a,b) \leq b \leq S(a,b)$$

and, furthermore it holds that

$$T(a,b) = 1 - S(1-a, 1-b)$$

which corresponds with the laws of De Morgan. Consequently, for each t-norm there exists a corresponding t-conorm. The t-norm min and t-conorm max presented by Zadeh in [5] are frequently used for their simple computability.

2.5.2 Aggregation Operations

As it has been mentioned, an aggregation operator over a finite number of fuzzy sets n —all defined over the same universe of discourse X— can generally be specified by means of a function $h : [0,1]^n \mapsto [0,1]$. Such a function h, to be a meaningful aggregation operator, minimally has to satisfy the following axioms:

- Axiom H1. $h(0,0,\ldots,0) = 0 \wedge h(1,1,\ldots,1) = 1$ (border condition).
- Axiom H2.

  $$\forall (a_1,a_2,\ldots,a_n),(b_1,b_2,\ldots,b_n) \in [0,1]^n, \forall_i \in \{1,2,\ldots,n\} :$$

  $$a_i \leq b_i \Rightarrow h(a_1,a_2,\ldots,a_n) \leq h(b_1,b_2,\ldots,b_n)$$

  (h is non-decreasing in all of its arguments).
- Axiom H3. h is a continuous function (continuity).

If an aggregation operator satisfies the following axioms H4 and H5, the operator is respectively said to be symmetric and idempotent.

- Axiom H4. For each permutation p of $\{1, 2, \ldots, n\}$ it must hold that

$$h(a_1, a_2, \ldots, a_n) = h(a_{p(1)}, a_{p(2)}, \ldots, a_{p(n)})$$

(h is a symmetric function in all of its arguments).

- Axiom H5. $\forall a \in [0, 1] : h(a, a, \ldots, a) = a$ (idempotency).

Examples of (classes of) aggregation operators are the so-called generalized averages [17] and the so-called ordered weighted averages (OWAs) [18].

The *generalized averages* are defined by

$$h(a_1, a_2, \ldots, a_n; \alpha) \triangleq \left(\frac{a_1^\alpha, a_2^\alpha, \ldots, a_n^\alpha}{n}\right)^{\frac{1}{\alpha}}$$

where $\alpha \in \mathbb{R}_0$ is the parameter that distinguishes the different aggregation operators and it must hold that $a_i \neq 0, i = 1, 2, \ldots, n$ if $\alpha < 0$.

The *ordered weighted averages* are defined by

$$h(a_1, a_2, \ldots, a_n; w) \triangleq w_1 b_1 + w_2 b_2 + \cdots + w_n b_n$$

where $w = (w_1, w_2, \ldots, w_n) \in [0, 1]^n$ is called the weighting factor. It must hold that $\sum_{i=1}^{n} w_i = 1$ and that for each $i \in \{1, 2, \ldots, n\}, b_i$ is the i^{th} largest element of $a_1, a_2, \ldots, a_n$, i.e. $(b_1, b_2, \ldots, b_n)$ is a permutation of $a_1, a_2, \ldots, a_n$ where the elements are ordered as follows:

$$\forall i, j \in 1, 2, \ldots, n : i < j \Rightarrow b_i \geq b_j.$$

For an state-of-the-art overview regarding these classes of aggregation operators the interested reader may refer to [19] as well as [20] where a .

2.5.3 Generalized Conjunction/Disjunction

The generalized conjunction/disjunction (GCD) operator is a continuous logic function that integrates conjunctive and disjunctive properties in a single function [21]. The GCD operator in $\mathbb{I}$ is a function

$$\begin{aligned} GCD : \quad & \mathbb{I}^n \to \mathbb{I} \\ & (x_1, \ldots, x_n) \mapsto x_1 \diamondsuit \ldots \diamondsuit x_n \end{aligned}$$

where $x_i \in [0, 1]$ for $i = 1, \ldots, n$ and $GCD \in [0, 1]$.

GCD includes two parameters: the *andness* and the *orness*. The *andness* α, denotes simultaneity and expresses the conjunction degree. Meanwhile, the *orness* ω, denotes replace-ability and expresses the disjunction degree. These parameters are complementary, i.e., $\alpha + \omega = 1$. Therefore, $\alpha = 1$ denotes full conjunction while $\omega = 1$ denotes full disjunction.

Although GCD can be implemented in several ways [22], within this dissertation the implementation based on the weighted power means (WPM) is presented considering its ability for interpreting preference logic [21] and modeling evaluation decisions [23].

GCD based on the weighted power means (WPM) is defined as follows:

$$x_1 \diamondsuit \ldots \diamondsuit x_n = (w_1 {x_1}^r + \cdots + w_n {x_n}^r)^{\frac{1}{r}}, \tag{2.2}$$

where w_i denotes the weight assigned to parameter x_i and the parameter r can be computed as a function of andness α using a suitable numerical approximation [23].

Table 2.1 includes, as a reference, the corresponding orness, andness and exponent r for 17 levels of GCD implemented using WPM. It could be noticed that symbols D and C correspond to full disjunction ($\omega = 1$), and full conjunction ($\alpha = 1$) respectively. Additionally, symbol A corresponds to the arithmetic mean where the conjunction and disjunction properties are balanced, i.e., $\alpha = \omega = 0.5$. Symbols starting with the letter D denote partial disjunction where the disjunction properties predominate, i.e., $\omega > 0.5$. Analogously, symbols starting with the letter C denote partial conjunction where the conjunction properties predominate, i.e., $\alpha > 0.5$.

Table 2.1: Aggregation operators for 17 levels of GCD implemented by WPM*.

Symbol	**Orness**(ω)	**Andness**(α)	**Exponent r**
D	1	0	$+\infty$
D++	0.9375	0.0625	20.63
D+	0.8750	0.1250	9.521
D+-	0.8125	0.1875	5.802
DA	0.7500	0.2500	3.929
D-+	0.6875	0.3125	2.792
D-	0.6250	0.3750	2.018
D–	0.5625	0.4375	1.449
A	0.5	0.5	1
C–	0.4375	0.5625	0.619
C-	0.3750	0.6250	0.261
C-+	0.3125	0.6875	-0.148
CA	0.2500	0.7500	-0.72
C+-	0.1875	0.8125	-1.655
C+	0.1250	0.8750	-3.510
C++	0.0625	0.9375	-9.06
C	0	1	$-\infty$

It is worth noting that other implementations of GCD may lead to different levels of andness and orness measures. For instance, the orness measures for two compound quasi-arithmetic mean aggregation operators is presented in [24]. Moreover, to obtain orness measures of well-known aggregation operations are emerging such as the one presented in [25] with regard to the OWAs.

2.5.3.1 GCD Verbalized Approach

The GCD verbalized approach [26] facilitates the selection of GCD aggregation operators by means of a multilevel overall importance scale. In this approach, a decision maker uses a scale with L levels from "lowest" to "highest" to specify the degree of overall importance for each attribute. Table 2.2 shows an overall importance scale of $L = 16$ levels [26].

Level (l)	**Overall importance**
16	Highest
15	*Slightly below highest*
14	Very high
13	*Slightly above high*
12	High
11	*Slightly below high*
10	Medium-high
9	*Slightly above medium*
8	Medium
...	...
0	Lowest

Table 2.2: Overall importance scale with $L = 16$ levels.

Besides the overall importance for each attribute, the decision maker should specify the selection of simultaneity or replace-ability among them. Herein, the selection of *simultaneity* refers to the fact that all the attributes need to be satisfied. In contrast, the selection of *replace-ability* means that a high aggregation value can be given by any of the attributes because one of them can replace the others.

On the one hand, the selection of simultaneity allows a decision maker to obtain the level of andness (denoted by α) as the mean of overall importance among the attributes given by

$$\alpha = \frac{(l_1 + \cdots + l_n)}{n \cdot L}, \; l_i \in [0, L] \tag{2.3}$$

where n corresponds to the number of attributes, l_i denotes the numerical value representing the level of each attribute i for $i = 1, \ldots, n$, and $L = 16$ corresponds to the maximum level on the overall importance scale.

On the other hand, the selection of replace-ability allows a decision maker to obtain the level of orness (denoted by ω) as the mean of overall importance among the attributes given by

$$\omega = \frac{(l_1 + \cdots + l_n)}{n \cdot L},\ l_i \in [0, L] \tag{2.4}$$

where n corresponds to the number of attributes, l_i denotes the numerical value representing the level of each attribute i for $i = 1, \ldots, n$, and $L = 16$ corresponds to the maximum level on the overall importance scale.

Recalling that GCD includes the parameter andness and its complement orness, in this way a GCD aggregation operator can be obtained.

For example, a decision maker may consider $n = 3$ attributes a_1, a_2 and a_3 with equal degree of importance denoted by 'medium' with level=8 in the overall importance scale, i.e. the level of these attributes correspond to $l_1 = l_2 = l_3 = 8$. In the case that, according to the decision maker, the attributes a_1, a_2 and a_3 need to be simultaneously satisfied, the andness α is given by

$$\begin{aligned}\alpha &= \frac{(l_1 + l_2 + l_3)}{n \cdot L}\\ &= \frac{(8 + 8 + 8)}{3 \cdot 16}\\ &= 0.5.\end{aligned}$$

The α-value previously obtained corresponds to the arithmetic mean, i.e., the GCD operator denoted by A. This GCD operator is obtained by scanning the column *Andness* (α) to look for the computed α-value in Table 2.1.

In the case that the attributes can be replaced by each other, according to a decision maker, the ω-value is computed by means of Equation 2.4. Thus, the GCD operator is obtained by scanning the column *Orness* (ω) in Table 2.1 when looking for the computed ω-value.

It is important to mention that when using Equations 2.3 and 2.4, a computed value that is not present in the Table 2.1 can be obtained. In this case, it is recommended to look for the closest computed α-value (or ω-value) to obtain the GCD operator.

Once the level of andness/orness have been obtained, it is necessary to map these into normalized weights $w_1 + \cdots + w_n = 1$. Although it is possible to use the GCD verbalized approach to compute the weights [26], in the remainder of this dissertation it is considered that these will also be given by the decision maker when obtaining the appropriate aggregator.

References

[1] Lotfi Zadeh. *Fuzzy logic, neural networks, and soft computing.* Communications of the ACM, 37(3):77–84, 1994.

[2] Francisco Herrera, Luis Martínez, and Pedro J. Sánchez. *Managing non-homogeneous information in group decision making.* European Journal of Operational Research, 166(1):115–132, 2005.

[3] Xia Chen, Hengjie Zhang, and Yucheng Dong. *The fusion process with heterogeneous preference structures in group decision making: A survey.* Information Fusion, 24:72–83, 2015.

[4] Michel Grabisch, Jean-Luc Marichal, Radko Mesiar, and Endre Pap. *Aggregation Functions (Encyclopedia of Mathematics and its Applications).* Cambridge University Press, New York, 1st edition, 2009.

[5] Lofti Zadeh. *Fuzzy Sets.* Information and control, 8(3):338–353, 1965.

[6] Witold Pedrycz, Petr Ekel, and Roberta Parreiras. *Fuzzy Multicriteria Decision-Making: Models, Methods and Applications.* John Wiley and Sons, Chichester, UK., 2011.

[7] George J. Klir and Bo Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice Hall, 1995.

[8] Patrice Perny and Marc Roubens. *Fuzzy preference modeling.* In Fuzzy sets in decision analysis, operations research and statistics, pages 3–30. Springer, 1998.

[9] Didier Dubois and Henry Prade. *Fundamentals of Fuzzy Sets (The Handbooks of Fuzzy Sets Volume 7).* Springer, 1st edition, 2000.

[10] Didier Dubois and Henry Prade. *The Three Semantics of Fuzzy Sets.* Fuzzy Sets and Systems, 90:141–150, 1997.

[11] Lotfi Zadeh. *The concept of a linguistic variable and its application to approximate reasoning-I.* Information Sciences, 8(3):199–249, January 1975.

[12] Lotfi Zadeh. *The concept of a linguistic variable and its application to approximate reasoning-II.* Information sciences, 8(4):301–357, 1975.

[13] Lotfi Zadeh. *The concept of a linguistic variable and its application to approximate reasoning-III.* Information sciences, 9(1):43–80, 1975.

[14] George A. Miller. *The magical number seven or minus two: Some limits on our capacity of processing information.* Psychol. Rev., 63:81–97, 1956.

[15] Gleb Beliakov, Tomasa Calvo, and Simon James. *On Penalty-Based Aggregation Functions and Consensus*, pages 23–40. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[16] Ildar Batyrshin and Okyay Kaynak. *Parametric classes of generalized conjunction and disjunction operations for fuzzy modeling.* IEEE Transactions on Fuzzy Systems, 7(5):586–596, 1999.

[17] Benito Vittorio Frosini. *Averages.* Cleup, 1987.

[18] Ronald R. Yager. *Families of OWA operators.* Fuzzy Sets and Systems, 59(2):125–148, oct 1993.

[19] Michel Grabisch, Jean-Luc Marichal, Radko Mesiar, and Endre Pap. *Aggregation functions: Means.* Information Sciences, 181(1):1–22, 2011.

[20] Ronald R. Yager and Naif Alajlan. *A generalized framework for mean aggregation: Toward the modeling of cognitive aspects.* Information Fusion, 17:65 – 73, 2014. Special Issue: Information fusion in consensus and decision making.

[21] Jozo Dujmović and Henrik Legind Larsen. *Generalized conjunction/disjunction.* International Journal of Approximate Reasoning, 46(3):423–446, December 2007.

[22] Jozo Dujmović. *Characteristic forms of generalized conjunction/ disjunction.* 2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence), pages 1075–1080, June 2008.

[23] Jozo Dujmović. *Continuous Preference Logic for System Evaluation.* IEEE Transactions on Fuzzy Systems, 15(6):1082–1099, December 2007.

[24] Xinwang Liu. *The orness measures for two compound quasi-arithmetic mean aggregation operators.* International Journal of Approximate Reasoning, 51(3):305 – 334, 2010.

[25] Amar Kishor, Amit K Singh, and Nikhil R Pal. *Orness measure of OWA operators: A new approach.* IEEE Transactions on Fuzzy Systems, 22(4):1039–1045, 2014.

[26] Jozo Dujmović. *Andness and orness as a mean of overall importance.* In Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on, pages 1–6. IEEE, 2012.

Chapter 3

Handling a Large Number of Opinions

Parts of this chapter were published in:

- Tapia-Rosero Ana, Bronselaer Antoon and De Tré Guy. **Similarity of membership functions: a shaped based approach**. In *Proceedings of the 4th International Joint Conference on Computational Intelligence*, 402-409. Barcelona, Spain, 2012.
- Tapia-Rosero Ana, Bronselaer Antoon and De Tré Guy. **A method based on shape-similarity for detecting similar opinions in group decision making**. *Information Sciences*, 258 (2014): 291-311.

3.1 Introduction

A decision-making problem may be solved by involving several persons in the process —like potential customers of a product, citizens of a country, or students in an educational institution. However, as elaborated in Chapter 1, the complexity of decision-making is related to the number of persons that are involved. Hence, when all the persons within a large group contribute to some extent to the final decision, this becomes a more complex task to pursue. In this dissertation, a relevant question that arises is: *How to simplify the complexity of a decision-making problem that involves a large number of persons in the decision?* The following example may provide a guide to find an answer to this question.

A decision on the stay of a product in the market. A company has to decide if a product stays in the market based on its "acceptable level of sales" (criterion). Here, each manager of the company may express what he/she understands to be an acceptable sales level.

In general, when different persons express their opinions (or preferences) three scenarios might be possible: (i) all the persons have a similar opinion; (ii) they all give a dissimilar opinion; or (iii) there are several groups of persons with similar opinions. Herein, by 'opinion' is meant a thought about the level of agreement over a specific criterion —like the *acceptable level of sales* of a product in the given example.

The first scenario may be considered as being of *low complexity* because all the persons share their agreement on the criterion. The second scenario may be considered as *highly complex* because the number of different opinions equals the number of persons involved in a decision. And the third scenario may be considered somewhere between the other two. Although these scenarios may be differentiated, this chapter only refers to the third scenario considering that the remaining scenarios are special cases where the level of complexity is the lowest or the highest respectively.

One strategy to reduce the complexity of the problem when handling a large number of opinions is to group them by similarity. In this way, given n opinions one might expect (after detecting similar opinions) m groups of similar opinions where $m \leq n$. Thus, a decision maker (e.g., the head of the company) could make a choice within a reduced set of opinions (Figure 3.1).

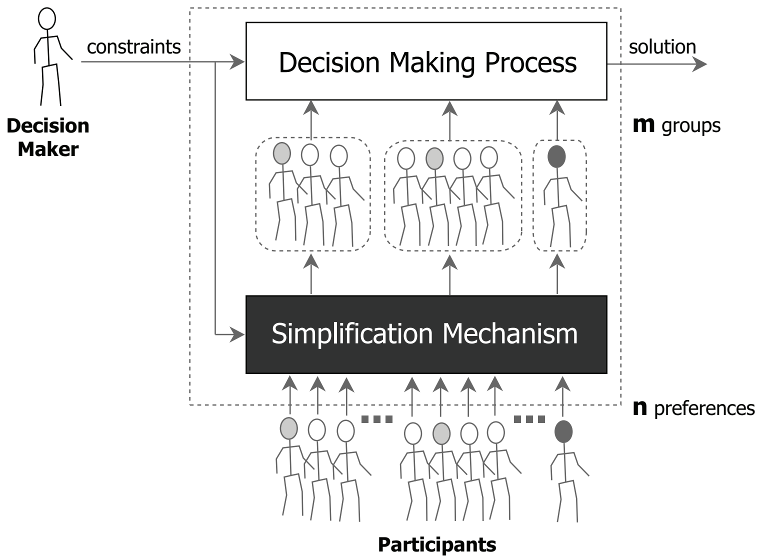


Figure 3.1: A strategy to handle a large number of opinions (or preferences) in order to reduce the complexity of decision making.

This study assumes that the similarity among opinions (i.e., using preference modeling) representing the same concept (e.g., the acceptable level of sales) may be detected by the similarity on the shape characteristics of their corresponding membership functions. For instance, two triangular membership functions expressing that the acceptable level of sales is around certain close values such as "around 45%" and "around 50%". Therefore, the purpose of this chapter is to detect similar opinions where opinions are unified by means of fuzzy sets in a common domain (Section 2.3.1), and the similarity focuses on a shaped based approach.

For that reason, this chapter describes a novel *shape-similarity detection method* which is feasible to be used in the presence of a large amount of membership functions. This method includes the so-called symbolic notation as a novel component. The symbolic notation is used to depict shape characteristics of membership functions facilitating the function comparisons considering those that are similarly shaped from a human's perspective (e.g., a triangular membership function and a trapezoidal membership function with a tiny core). In this way, the main contribution of this chapter is a shape-similarity detection method [1] which may assist a decision maker managing the complexity in a decision when handling a large number of opinions (e.g., when a community is involved).

The structure of the chapter is as follows: Section 3.2 provides an overview of related work with regard to fuzzy similarity. Section 3.3 provides details on the shape-similarity detection method including its general architecture and details on its component phases. Section 3.4 presents an example to illustrate the applicability of the proposed method. Section 3.5 presents the answer to the introductory question and the contributions of this chapter.

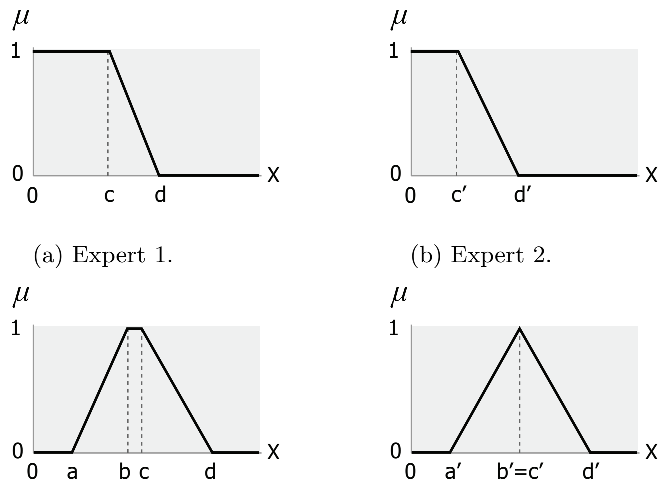
3.2 Related Work

This chapter assumes, as a starting point, that two membership functions are considered to be similar if they have a "similar" shape. One remark is that similarly shaped membership functions should represent the same concept as expressed by the expert (or person involved in the decision).

For instance, Figure 3.2 shows four expert opinions represented by membership functions expressing "the usefulness level of a product". Despite that the X-axis values are different, opinions of experts 1 (Figure 3.2a) and 2 (Figure 3.2b) could be considered similar according to their shape, likewise the opinions of experts 3 (Figure 3.2c) and 4 (Figure 3.2d). Hence, the aim of this section is to review some related work on fuzzy similarity, where fuzzy sets are used for criteria definition (Section 2.3) in a decision-making context.

3.2.1 Fuzzy Similarity

Various similarity measures exist in order to compare fuzzy sets, which are based on one of the following ideas or on a combination thereof: (i) similarity relations, (ii) distance among fuzzy sets, and (iii) set-theoretic operations. These are presented next considering that most of the similarity measures are based on them.



(a) Expert 1. (b) Expert 2.

(c) Expert 3. (d) Expert 4.

Figure 3.2: Expert opinions represented by membership functions. Here, the opinions of experts 1 and 2 could be considered similar according to their shape, likewise the opinions of experts 3 and 4.

3.2.1.1 Similarity Relations

The definition of a similarity relation was introduced by [2] as an extension of the equivalence relation concept as follows.

Definition 3.1 (Similarity Relation [2])
A similarity relation S is a fuzzy relation which is reflexive, symmetric, and transitive.

According to this definition the following properties hold:

- $\mu_S(x,x) = 1$ (reflexivity),
- $\mu_S(x,y) = \mu_S(y,x)$ (symmetry), and
- $\mu_S(x,z) \geq \mu_S(x,y) \wedge \mu_S(y,z)$ (transitivity)

where x, y, z are elements of a set X, $\mu_S(x,y)$ denote the grade of membership of the ordered pair (x,y) in a similarity relation S and $\wedge$ represents the t-norm *min* (Section 2.5.1).

Thus, a similarity relation can be also called a fuzzy equivalence relation which can be interpreted in two different ways [3] as follows. The first interpretation considers that the elements of a fuzzy set can be grouped in sets where its members are similar to some specified degree. Here, it is worth noting that when this degree equals 1, the grouping corresponds to an equivalence class. The second interpretation considers “the degree of similarity that the elements of X have to some specified element $x \in X$" [3]. In this case, a fuzzy set defines a similarity class where the degree of similarity of a particular element corresponds to its membership grade. Analogously, if all the elements within a class are similar to the specified element x to the degree of 1 and similar to the degree of 0 to all the elements outside this class, then the grouping becomes an equivalence class.

3.2.1.2 Distance Among Fuzzy Sets

> The notion of a distance, $d(x, y)$, between objects x and y has long been used in many contexts as a measure of similarity or dissimilarity between elements of a set [2].

There are different measures for the distance $d(x, y)$, between objects x and y, based on known metrics such as Hausdorff [4], Hamming [5], Manhattan [6] and Euclidean [7] distances. Some of these distance based similarity measures are considered theoretical approaches because they suppose the existence of an "ideal" fuzzy set. However, most of the decision-making problems, including those with multiple experts, do not have a given "ideal" solution. Usually, in a decision-making problem, one is looking for a resulting fuzzy set that adequately represents a concept in a given context.

An experimental study presented in [8] shows that several distance functions produce different results in most of the carried out group decision-making problems. It was also observed that in most cases with a greater number of experts, there is a trend of results being identified as significantly different.

To the best of the author's knowledge none of the distance based proposals make a distinction with regard to the shape of the membership functions. However, in [9] is proposed a vector similarity measure with two components: the similarity in shape and the proximity. In that study the shape-similarity in membership functions does not look for a particular shape, but uses a centroid alignment technique for comparisons. Meanwhile the proximity component is calculated using the Euclidean distance between their centroids.

3.2.1.3 Set-theoretic Operations

Set-theoretic operations for fuzzy sets allow for combining several membership functions (cf. Section 2.5). There are several strategies to combine membership functions including Zadeh's basic operations among fuzzy sets for union (max) and intersection (min) among other definitions based on triangular norms (t-norms) and triangular conorms (t-conorms) —e.g., the product and the probabilistic sum, the Lukasiewicz t-norm and t-conorm [10].

Several similarity measures are based on set-theoretic operations such as Tversky [11], Dice [12] and Jaccard indexes [13]. Most of these similarity measures have special applications related to the nature of the involved sets, i.e., if one set is included in the other or if sets overlap.

Within the scope of this dissertation, it is feasible to consider that two fuzzy sets are similar to some extent even if they do not overlap. An example borrowed from [14] illustrates this as follows: *"Two experts assign different supports to an alternative under a criterion ranging from 0.78 to 0.80 and 0.81 to 0.83, respectively"*. For the purpose of illustration, this example is depicted in Figure 3.3 where two fuzzy sets have disjoint supports but represent two close opinions. One disadvantage on using set-theoretic operations is that similarity between these fuzzy sets equals zero.

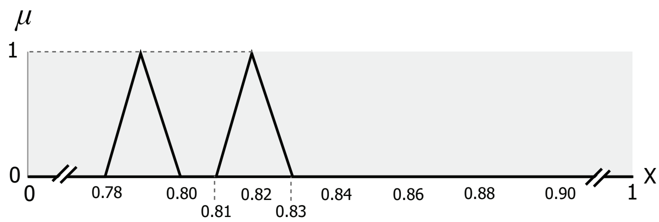


Figure 3.3: Two fuzzy sets with disjoint supports representing close opinions.

A state of the art of similarity measures for intuitionistic fuzzy sets is presented in [15]. And, a more recent book [16] presents not only a very comprehensive set of similarity measures for intuitionistic fuzzy sets, but a set of distance measures for this type of fuzzy sets. Additionally, in [17] various distance and similarity measures for hesitant fuzzy sets are presented. Another work [18] presents the Hamming distance, Euclidean distance, Hausdorff distance and a generalized distance between interval-valued hesitant fuzzy sets.

An interesting research is presented in [19] where many measures of similarity among fuzzy sets are compared in a behavioral experiment. In this work, all the measures performed well when categorizing membership functions as similar or dissimilar, but there is a difference in their performance when trying to distinguish grades of similarity or dissimilarity. Additionally, in this work it has been pointed out that "the best measures were ones that focus on only one 'slice' of the membership function".

3.2.2 Shape-Similarity Measure

The similarity measure presented later in this chapter is constructed using a shape based approach (Section 3.3.2), i.e., two membership functions are considered to be similar if they have a "similar" shape while reflecting similar expert opinions (using preference modeling).

An initial symbolic notation was presented in [20] to represent the shape of membership functions. This work introduced a straightforward similarity measure which performs symbolic-notation comparisons. Then, the similarity measure was extended in [1] by including a linguistic component which allows for making better comparisons while keeping simple notations.

Next section presents the shape-similarity detection method.

3.3 Shape-Similarity Detection Method

The *shape-similarity detection method* aims to reduce the complexity of a decision-making problem related to the number of opinions given by persons that are involved in a decision. The idea is that this method receives n opinions as inputs and provides m groups of similar opinions as outputs, where $m \leq n$. Figure 3.4 presents a simplified diagram of the shape-similarity detection method.

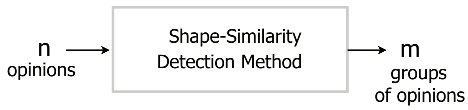


Figure 3.4: Simplified diagram of the shape-similarity detection method.

Before proceeding with the shape-similarity detection method, it is important to express herein what is meant by 'shape-similarity' with regard to handling a large number of opinions. First, one can recall from the previous chapter that the definition of a criterion, and hence an opinion, is represented by means of a piecewise membership function (i.e., a triangular or a trapezoidal membership function). Then, it can be expressed that 'shape-similarity' refers to the similarity based on the shape characteristics of membership functions used to represent the opinions on a specific criterion.

To describe the shape-similarity detection method the following phases are detailed: (i) obtaining symbolic notations; (ii) computing shape-similarities; and (iii) grouping by shape-similarity. Figure 3.5 depicts the general architecture of the shape-similarity detection method.

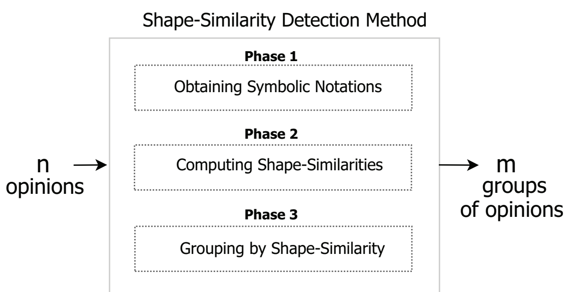


Figure 3.5: General architecture of the shape-similarity detection method.

3.3.1 Obtaining Shape-Symbolic Notations

Bearing in mind the presence of a large number of membership functions, a proper notation of these may significantly reduce the complexity of the problem of shape-similarity detection when handling a large number of opinions. Therefore, for each normalized membership function (see Definition 2.2) a *shape-symbolic notation* is obtained.

The *shape-symbolic notation* of a membership function has two components, namely a *shape-string* and a *feature-string*. The *shape-string* component denotes the shape characteristics of the membership function, i.e. slopes and levels of preference. The *feature-string* component expresses the relative length of the shape characteristics on their X-axis segments, i.e. core segments, left and right spreads.

3.3.1.1 Shape-String

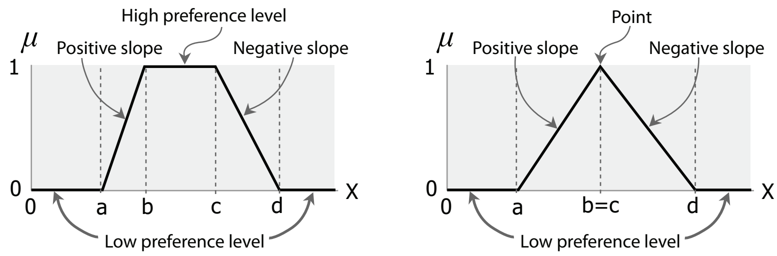
In this study, trapezoidal membership functions are used to represent opinions. The presence of segments in a trapezium shape can be graphically identified. These segments are based on the intervals that are present in the membership function definition (Equation 2.1), and each of them belongs to one of the following categories:

- Positive slope $\{+\}$
- Negative slope $\{-\}$
- Low preference level $\{0\}$
- High preference level $\{1\}$
- Point $\{L, I, H\}$, where a letter denotes a *low, intermediate* or *high* membership value respectively. Herein, L corresponds to 0, H corresponds to 1 and I corresponds to a membership value in]0,1[.

  Although there is an equivalence between a *high (or low) preference level* and a *high (or low)* membership value regarding the *point* category, the segments belonging to these categories are different. The main difference between these segments is that a *(high or low) preference level* is graphically identified as a horizontal segment, while the *(H or L) point* depicts as suggested by its name a point.

As an illustration, Figure 3.6 shows a trapezoidal and a triangular membership function with their corresponding segment categories —here, one can recall that a triangular membership function is a special case of a trapezoidal function where $b = c$ (denoting a typical value). Here, it can be noticed that the typical value of the triangular membership function uses the *point* category, while the trapezoidal membership function uses the *high preference level* category.

Formally, $S^{\text{category}} = \{+,\ -,\ 0,\ 1,\ L,\ I,\ H\}$ is considered to be the set of symbols that is used to represent the category of a segment in a membership function, and a *shape-string* is defined as follows:



(a) (b)

Figure 3.6: (a) Segment categories in a trapezoidal membership function. (b) Segment categories in a triangular membership function.

Definition 3.2 (Shape-String)
A shape-string *is a sequence of symbols that denotes the shape of a membership function according to their order on the X-axis, where each symbol is an element of the* S^{category} *set.*

An option to generate the *shape-string* is using a context-free grammar $G(N, T, S_0, P)$, where N is the set of non-terminal symbols, T is the set of terminal symbols, S_0 is the starting symbol as follows

`N` = {`<slope>,<preference level>,<point>,<segment>,<shape-string>`};
`T` = {`+,-,0,1,L,I,H`};
S_0 = `<shape-string>`;
and `P` is the set of the following production rules:
`<slope>::= + | -`
`<preference level>::= 0 | 1`
`<point>::= L | I | H`
`<segment>::= <slope> | <preference level> | <point>`
`<shape-string>::= <segment> | <shape-string><segment>`

Additionally, it is possible to define a linguistic variable (see Definition 2.3) to formally represent the *point* category as follows:

- V = point, is the variable name;
- $T(V) = \{$low, intermediate, high$\}$, is the term set;
- $U = [0, 1]$, is the universe of discourse;
- $M(\text{low}) = L$, $M(\text{intermediate}) = I$ and $M(\text{high}) = H$ are the semantic rules that assigns a label to each linguistic term. In this case, the labels L, I and H can be fuzzy sets represented by membership functions μ_{low}, $\mu_{\text{intermediate}}$ and μ_{high} respectively.

To obtain a *shape-string* for a given membership function an algorithm is further provided (Algorithm 3.1), but first the general idea may be described as follows.

The *shape-string* of a membership function is built by concatenating the symbols that represent each segment taking into account their order on the X-axis. For instance, Figure 3.6 depicts two membership functions where the shape-strings "0+1−0" and "0+H−0" represent the trapezoidal and triangular membership functions respectively. In this example, it could be noticed that the shape-string of the triangular membership function uses the letter H which denotes a *high* membership value within the *point* category.

Algorithm 3.1 shows the steps to obtain the *shape-string* given a membership function μ and a sorted-list P. Herein, P is a sorted-list of parameters that is used as a generalized representation of the membership function μ —i.e., its parameters correspond to the X-axis coordinates. Figure 3.7 provides an illustration of a membership function represented by a sorted-list of parameters $P = [s, a, b, c, d, e]$. It could be noticed that in addition to the parameters a, b, c, d used to represent a trapezoidal membership function (Section 2.2.1). parameters s and e has been included in the sorted-list P. Here, parameters s and e allow for representing the domain limits of membership function μ.

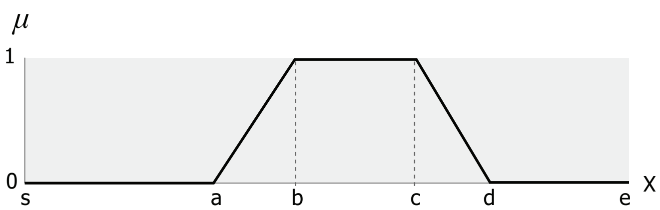


Figure 3.7: A trapezoidal membership function represented by a sorted-list of parameters $P = [s, a, b, c, d, e]$.

The algorithm *getShapeString* iteratively compares two consecutive parameters denoted as P_i and P_{i+1} (line 5) or their corresponding membership degrees (lines 14 and 17) to detect the category of a segment. From line 6 to line 24 the method assigns to the *symbol* variable: a character among a letter {L, I, H} to denote a *point*, a sign {+,−} to denote a *slope*, and a value {0, 1} to represent the *preference level* on segments without a slope. Each of these characters will be linked together into the string *sStr* through the *Concatenate* function (line 25). Finally, the string *sStr* representing the *shape-string* is returned.

In this way, a *shape-string* could represent different kinds of piecewise membership functions using the S^{category} set as described before. Nevertheless, this method might be extended to represent periodical functions and other special cases not considered within the scope of this study.

Algorithm 3.1 getShapeString

Require: membership-function μ
Require: sorted-list P
Ensure: string $sStr$

```
l ← length(P)
symbol ← “”
sStr ← “”
for i = 1 to l − 1 do
  if P[i] = P[i + 1]                          //Detection of point category
    if μ(P[i]) = 1
      symbol ←“H”                             //H corresponds to 1
    else
      if μ(P[i]) = 0
        symbol ←“L”                           //L corresponds to 0
      else
        symbol ←“I”     //I corresponds to a value in the range of ]0,1[
  else
    if μ(P[i + 1]) > μ(P[i])                  //Detection of positive slope
      symbol ←“+”
    else
      if μ(P[i + 1]) < μ(P[i])                //Detection of negative slope
        symbol ←“−”
      else
        if μ(P[i]) = 1                        //Detection of high preference level
          symbol ←“1”
        else
          if μ(P[i]) = 0                      //Detection of low preference level
            symbol ←“0”
  sStr ← Concatenate(sStr, symbol)
return sStr
```

3.3.1.2 Feature-String

In this study, the *feature-string* facilitates expressing relative length approximations of the segments on the X-axis of a membership function by means of linguistic terms (Section 2.4). Herein, each linguistic term represents the relative length of a shape feature of a membership function (i.e., core, left and right spreads) with regard to the overall length of the membership function. As an illustration, Figure 3.8 shows a trapezoidal membership function with its corresponding linguistic terms.

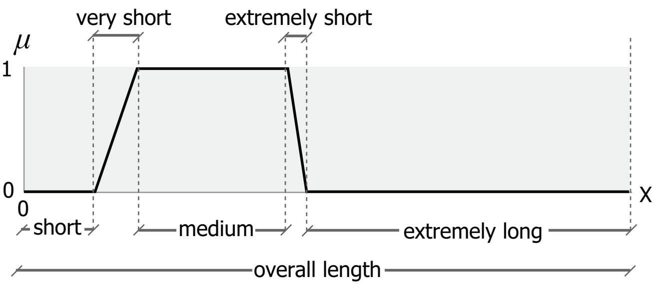


Figure 3.8: Length approximations of the segments of a membership function.

Here, it is important to mention that even though (theoretically) it is possible to have an unlimited domain for a membership function, i.e. $]-\infty,+\infty[$, for implementation purposes the domain has been limited to $[min, max]$.

The *feature-string* is obtained considering that the overall-segment length of a membership function (see Definition 3.3) is used as a standard when obtaining the linguistic terms representing the relative length of each of its segments. Once all the relative lengths are obtained the *feature-string* is built by concatenating the label that represents each linguistic term taking into account their order on the X-axis.

Formally, S^{length} is considered to be the linguistic term set that is used to represent the relative length of a segment compared to the sum of all segments in a membership function. Figure 3.9 depicts a set of linguistic terms expressing lengths ranging from “extremely short” to “extremely long”.

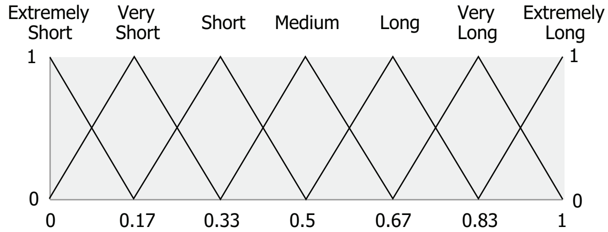


Figure 3.9: Linguistic term set S^{length} and its semantics.

To obtain the *feature-string* for a membership function μ an algorithm (Algorithm 3.2) that uses the following definitions is provided:

Definition 3.3 (Overall-Segment Length)
The overall-segment length *in a membership function μ is the sum of all segments' lengths on the X-axis.*

The *overall-segment length* is obtained by Equation 3.1:

$$osl(\mu) = \sum_{i=1}^{N} L_i(\mu). \tag{3.1}$$

Hereby, $L_i(\mu)$ denotes the length value of segment i on the X-axis of membership function μ. N is the number of segments of the function μ.

Definition 3.4 (Feature-String)
A feature-string *is a sequence of linguistic terms that denotes the relative length of each segment of a membership function μ compared to its* overall-segment length.

The relative length $r(i)$ on segment i is computed by:

$$r(i) = \frac{L_i(\mu)}{osl(\mu)}. \tag{3.2}$$

Hereby, $L_i(\mu)$ denotes the length value of segment i on the X-axis and $osl(\mu)$ is the overall-segment length of membership function μ.

As a next step the linguistic term that most adequately represents the relative length between segment i and the *overall-segment* of the corresponding membership function μ should be selected from the linguistic term-set S^{length}. Considering that two consecutive linguistic terms may be associated with a relative length r, this study will use the linguistic term with the highest membership grade for the relative length r to keep the shape-symbolic notation simple. However, other approaches[1] may be used and are subject to further study.

For instance, one may consider that the relative length r (obtained using Equation 3.2) is associated with the consecutive linguistic terms *very short* and *short* as shown in Figure 3.10. In this case, the membership grade of r for the linguistic term *very short* is $\mu_{\text{very-short}}$. Analogously, the membership grade of r for the linguistic term *short* is μ_{short}. So, the linguistic term *short* is selected to represent the relative length r in the case of the example —considering that $\mu_{\text{short}}(r) > \mu_{\text{very-short}}(r)$.

Next, Algorithm 3.2 shows how to obtain the *feature-string* given a linguistic-term set S and a sorted-list P that contains the membership function parameters including its domain limits $[s, e]$ (see Figure 3.7)—Despite the fact that

[1]The *2-tuple fuzzy linguistic representation model* for computing with words [21] may improve the accuracy, and the use of *hesitant fuzzy linguistic term sets* [22] can manage more than one linguistic term and hence may increase the flexibility within this case.

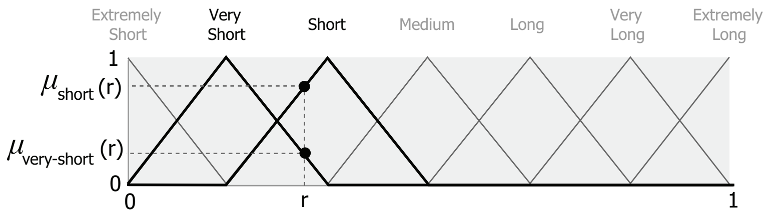


Figure 3.10: Linguistic terms *very short* and *short* associated to relative length r. Here, the linguistic term *short* is selected to represent r considering that $\mu_{\text{short}}(r) > \mu_{\text{very-short}}(r)$.

from a theoretical point of view, it is possible to have an unlimited domain for the membership function $]-\infty,+\infty[$, for implementation purposes the domain has been limited to $[min, max]$.

The algorithm assigns the length of sorted-list P to the variable *size* (line 3). Lines 4 and 5 obtain the *overall-segment length* according to Equation 3.1. The algorithm iteratively calculates the *segment length* between two consecutive parameters P_{i+1} and P_i (line 7). To obtain the *relative length* of the segment, it divides the *segment length* to the *overall-segment length* as indicated in Equation 3.2 (line 8). Next, it obtains the associated linguistic term (e.g., the linguistic term with the highest degree of membership) for each segment (line 9) through the *getLinguisticTerm* function. Finally, the *feature-string* is concatenated with the obtained linguistic term (line 10). Here, it will be necessary to use a delimiting character (e.g., hyphen, comma, etc.) to distinguish the linguistic term that corresponds to each segment, thus we consider that the *Concatenate* function (line 10) handles to include a string separator among linguistic terms. Finally, the string *fStr* representing the *feature-string* is returned (line 11).

For illustration purposes, within this dissertation, the function *getLinguisticTerm* will return the linguistic term with the highest degree of membership and the *Concatenate* function will use a hyphen as a string separator.

When the *shape-string* and the *feature-string* are obtained as indicated, this results in a *shape-symbolic notation* for a given membership function which is defined as follows:

Definition 3.5 (Shape-Symbolic Notation)
A shape-symbolic notation *is a representation of a membership function as a pair* $(sStr, fStr)$*, where* $sStr$ *corresponds to the* shape-string *denoting its shape characteristics; and* $fStr$ *corresponds to the* feature-string *expressing the relative length of its shape characteristics by means of linguistic terms.*

For instance, the shape-symbolic notation of the membership function shown in Figure 3.8 is ("0+1-0","S-VS-M-ES-EL").

In this way, phase 1 of the *shape-similarity detection method* results in a *shape-symbolic notation* for each available membership function in order to

Algorithm 3.2 getFeatureString

Require: linguistic-term-set S
Require: sorted-list P
Ensure: string $fStr$

$fStr \leftarrow$ “”
$osl \leftarrow 0$
$size \leftarrow length(P)$
for $i = 1$ to $size - 1$ **do**
    $osl \leftarrow osl + (P[i+1] - P[i])$ //According to Equation 3.1
for $i = 1$ to $size - 1$ **do**
    $segmentLength \leftarrow P[i+1] - P[i]$
    $relativeLength \leftarrow segmentLength/osl$
    $lTerm \leftarrow getLinguisticTerm(relativeLength, S)$
    $fStr \leftarrow Concatenate(fStr, lTerm)$ //Including a string separator
return $fStr$

facilitate the comparison among membership functions while reducing the complexity of the problem when handling a large number of them —where each membership function represents an opinion.

3.3.2 Computing Shape-Similarities

This phase computes a novel *shape-similarity measure* between two shape-symbolic notations by detecting differences when comparing their segments, and these differences are taken into account for penalty purposes. In this way, the *shape-similarity measure* provides a value in the unit interval where 0 denotes no similarity, 1 denotes full similarity and other intermediate values denote a partial similarity between two *shape-symbolic notations*. The latter may correspond to the case where two membership functions have partially matching shapes, and a technique to measure the difference between their corresponding notations is needed.

To measure the difference between two strings a technique based on the Levenshtein distance can be used. The Levenshtein distance calculates the minimum number of edit operations C_e such as insert, delete, or replace to transform one string into the other [23], where each edit operation is associated with a transform cost. Within this dissertation, a similar technique is given to measure the difference between two shape-symbolic notations. That is to say, to calculate the number of edit operations to transform one shape-symbolic notation into the other taking into account that shape-symbolic notations have two components —i.e., a shape-string and a feature-string. In this case, each edit operation is associated with a transform cost which is expressed in *cost-units*.

The formalization of an adapted version of the Levenshtein distance between two symbolic-notations requires the following definitions:

Definition 3.6 (Symbolic-Character)
A symbolic-character *is a representation of a segment in a membership function as a pair $\langle t, r \rangle$ with $t \in S^{\text{category}}$ and $r \in S^{\text{length}}$, where t represents the category of the segment and r denotes its relative length by means of a linguistic term.*

For example, Figure 3.11 shows a membership function with shape-string "0+1" and feature-string "S-VS-EL". This membership function has three segments and hence three symbolic-characters annotated as $\langle 0, S\rangle$, $\langle +, VS\rangle$ and $\langle 1, EL\rangle$.

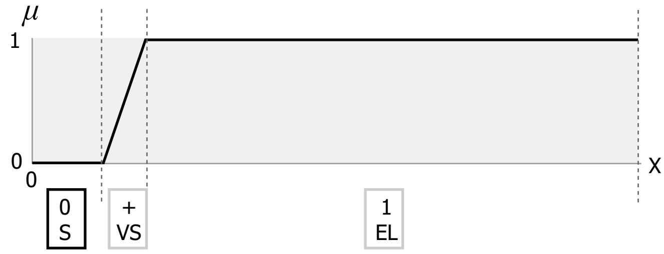


Figure 3.11: Representation of a *symbolic-character*. Here, the symbolic character $\langle 0, S\rangle$ appears highlighted.

At this point, it is possible that a *shape-symbolic notation* may also be denoted as a sequence of shape-symbolic characters. Moreover, the length of a shape-symbolic notation n denoted by length(n) corresponds to the number of symbolic characters contained in the shape-symbolic notation. For instance, if $n =$ "$\langle 0, S\rangle\langle +, VS\rangle\langle 1, EL\rangle$" then length($n$) = 3.

Definition 3.7 (Insert-Cost)
The insert-cost C_i *is the cost of inserting a symbolic-character into a shape-symbolic notation.*

The insert-cost of a symbolic-character includes one cost-unit due to the insertion of a character representing a shape characteristic (e.g., a slope), plus a number of cost-units representing the insertion of a linguistic term representing the length of the shape characteristic. In this way, it could be noticed that the number of cost-units associated to the linguistic component may vary depending on the linguistic term-set S^{length} that is used. For instance, within this study S^{length} ranges from "extremely short" to "extremely long" (see Figure 3.9). So, one may consider one cost-unit representing the minimum insertion cost associated to the linguistic term "extremely short" and seven cost-units representing the maximum insertion cost associated to the linguistic term "extremely long". However, for generalization purposes, in this dissertation it is considered that the insert-cost of a symbolic-character into a symbolic notation corresponds to: one cost-unit for inserting a character representing the shape, plus the maximum cost that is possible when inserting the feature-component. This maximum cost is represented by the cardinality of the linguistic term-set S^{length} denoted as $\mid S^{\text{length}} \mid$.

Definition 3.8 (Delete-Cost)
The delete-cost C_d *is the cost of deleting a symbolic-character from a shape-symbolic notation.*

By analogy to the insert-cost, the delete-cost of a symbolic-character from a shape-symbolic notation corresponds to one cost-unit, for deleting a character representing the shape, plus the maximum possible cost of deleting the feature-component —i.e., the cardinality $\mid S^{\text{length}} \mid$ of the linguistic term-set S^{length}.

Definition 3.9 (Replace-Cost)
The replace-cost C_r *is the cost of replacing a symbolic-character in a shape-symbolic notation.*

A replace operation is used when a symbolic-character changes. On the one hand, it is possible to change the character representing the shape of a segment. In this case, the replace-cost corresponds to one cost-unit. Here, it is also possible to apply a replace operation on characters representing equivalent shapes —e.g., a high preference level denoted as 1 and a point with the highest preference denoted as H. On the other hand, it is possible to replace the linguistic term representing the relative length of a segment. In this case, the replace-cost corresponds to the number of cost-units associated to the number of positions between the original linguistic term and the one that will be replaced. In the case that both components of the symbolic-character change, the replace-cost will be the aggregation of the aforementioned costs.

The following example illustrates the impact of edit operations when the shape-symbolic notation n_1 depicted in Figure 3.11 is transformed into the shape-symbolic notation n_2 depicted in Figure 3.12, where

$$
\begin{aligned}
n_1 &= \langle 0, S\rangle\langle +, VS\rangle\langle 1, EL\rangle \\
n_2 &= \langle 0, S\rangle\langle +, VS\rangle\langle 1, M\rangle\langle -, S\rangle\langle 0, M\rangle.
\end{aligned}
$$

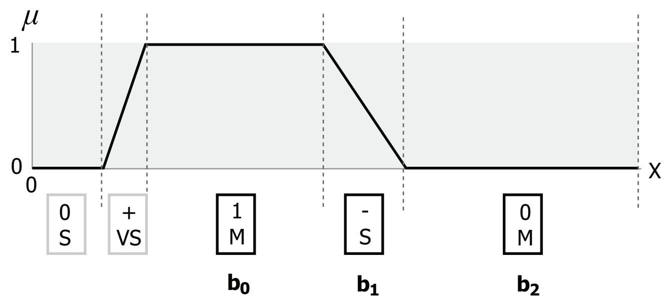


Figure 3.12: Replacement of symbolic-character $b_0 = \langle 1, M\rangle$ and insertion of symbolic-characters $b_1 = \langle -, S\rangle$ and $b_2 = \langle 0, M\rangle$ in shape-symbolic notation n_1 =("0+1","S-VS-EL").

Replacing a symbolic-character. The replace-cost from symbolic-character $\langle 1, EL\rangle$ to symbolic-character $\langle 1, M\rangle$, represented by $Cr_{\langle 1,EL\rangle \frown \langle 1,M\rangle}$, includes three cost-units representing the change of relative-length from "extremely long" to "medium" —i.e., moving three positions considering that the linguistic terms "long" and "very long" are between them. It could be noticed that in this case the shape component stays unchanged.

Inserting symbolic-characters. For this purpose, two symbolic-characters $b_1 = \langle -, S\rangle$ and $b_2 = \langle 0, M\rangle$ should be inserted at the end of shape-symbolic notation n_1. Thus, the total insert-cost C_i corresponds to the sum of the insert-costs of these symbolic-characters as follows:

$$C_i = C_{i_{\langle -,S\rangle}} + C_{i_{\langle 0,M\rangle}}$$

The insert-cost of a symbolic-character into a symbolic notation corresponds to one cost-unit, for inserting a character representing the shape, plus the maximum cost that is possible when inserting the feature-component, i.e. the cardinality of the linguistic term-set given by $\mid S^{\text{length}} \mid = 7$. Therefore, the insert-cost of symbolic-characters b_1 and b_2 is $1+ \mid S^{\text{length}} \mid = 8$ cost-units for each symbolic character.

Once the costs associated with converting one shape-symbolic notation into another are obtained, a *shape-similarity measure* can be defined as follows:

Definition 3.10 (Shape-Similarity Measure)
A shape-similarity measure *is a value of the unit interval* ($\in [0, 1]$) *denoting the similarity between two shape-symbolic notations n_1 and n_2.*

The shape-similarity measure *is calculated as follows:*

$$S(n_1, n_2) = 1 - p(n_1, n_2), \tag{3.3}$$

where $p(n_1, n_2)$ is the notation-penalty between notations n_1 and n_2.

The *notation-penalty* $p(n_1, n_2)$ for shape-symbolic notations n_1 and n_2 is calculated as follows:

$$p(n_1, n_2) = \frac{C_e(n_1, n_2)}{e_{max}(n_1, n_2)}. \tag{3.4}$$

Hereby, C_e is the cost of edit operations (insert, delete and replace) on shape-symbolic notations n_1 and n_2; and e_{max} is the maximum cost of possible edit operations between shape-symbolic notations n_1 and n_2 given by:

$$e_{max}(n_1, n_2) = \max(\text{length}(n_1), \text{length}(n_2)) \cdot (1+ \mid S^{\text{length}} \mid), \tag{3.5}$$

where $\max(\text{length}(n_1), \text{length}(n_2))$ is the maximum length between shape-symbolic notations n_1 and n_2, and $\mid S^{\text{length}} \mid$ is the cardinality of the linguistic term-set S^{length}.

The presented *shape-similarity measure* provides a value in the unit interval, where 0 denotes no similarity, 1 denotes full similarity between two *shape-symbolic notations*, and intermediate values denote partial similarity. It could be noticed that the *shape-similarity measure* satisfies the reflexivity and symmetry properties because:

- $C_e(n_1, n_2) = C_e(n_2, n_1)$, where C_e is the number of edit operations on shape-symbolic notations n_1 and n_2.
- $e_{max}(n_1, n_2) = e_{max}(n_2, n_1)$, where e_{max} is the maximum number of possible edit operations between shape-symbolic notations n_1 and n_2.

In this way, the computation of the similarity among a large amount of membership functions (represented by their corresponding shape-symbolic notations) uses a shape-based approach. Therefore, the next phase of the *shape-similarity detection method* looks for grouping similarly shaped membership functions.

3.3.3 Grouping by Shape-Similarity

This phase aims to provide m groups of similarly shaped membership functions based on the values obtained when computing their similarities. Thus, two approaches can be followed. One of them consists in that the decision maker fixes the m value according to his/her preferences (constraint). Another approach consists in that the decision maker establishes a value (threshold) that should be exceeded by the computed similarity. In this case, two membership functions belong to the same group when their similarity (measure) exceeds the established threshold. It is worth to mention that the membership functions contained in a group satisfy the commutative, distributive and associative properties.

Considering that the *shape-similarity measure* previously presented corresponds to a value in the unit interval, it is possible to obtain: (i) groups that consists of a single membership function denoting that these membership functions are dissimilar with others; (ii) a group that contains all the available membership functions denoting that all of them are similar; or (iii) groups with a varied number of membership functions denoting that all the membership functions that belong to the same group are considered to be similar to some extent.

To facilitate forming the groups of similarly shaped membership functions a *similarity matrix* is used.

Definition 3.11 (Similarity Matrix)
A similarity matrix $M_{k \times k}$ *for a group of* k *membership functions represented by their corresponding* shape-symbolic notations $n_1, n_2, \ldots, n_k$ *indicates the* shape-similarity measure S_{ij} *between notations* n_i *and* n_j, *where* $i = 1, \ldots, k$; $j = 1, \ldots, k$; *and* $0 \leq S_{ij} \leq 1$.

$$M = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{21} & S_{22} & \cdots & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{k1} & S_{k2} & \cdots & S_{kk} \end{pmatrix}$$

For grouping purposes, there are several strategies that could be used. One of them is to consider that each shape-symbolic notation initially corresponds to one group and hence, there will be as much groups as there are shape-symbolic notations. In this case, two groups are iteratively merged based on the highest shape-similarity value available in the similarity matrix. And, the grouping stops when the highest similarity is considered too low according to a previously determined threshold. This technique is called *agglomerative hierarchical clustering* with the single linkage rule as merge criterion [24], and hence hereafter the term *cluster* may be also used to represent a group of similarly shaped membership functions.

The hierarchical clustering algorithm may be graphically represented by the so-called dendrogram where the initial nodes correspond to shape-symbolic notations, the internal nodes correspond to a cluster containing similarly-shaped membership functions and the top nodes correspond to clusters that have been merged into a larger cluster.

A sample dendrogram is depicted in Figure 3.13 where the initial nodes correspond to shape-symbolic notations n_1, n_2, n_3, n_4 and n_5. Here, depending on the threshold value, the following clusters are formed: (i) when the threshold is τ_1, there are 5 clusters, each containing only one shape-symbolic notation; (ii) when the threshold is τ_2, the clusters $\{n_1, n_2\}$,$\{n_3, n_4\}$ and $\{n_5\}$ can be identified; (iii) when the threshold is τ_3, the clusters $\{n_1, n_2, n_3, n_4\}$ and $\{n_5\}$ are formed; and (iv) when the threshold is τ_4, all the shape-symbolic notations are contained in a single group.

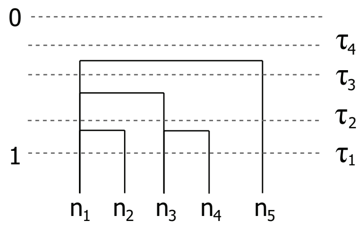


Figure 3.13: A dendrogram where the initial nodes are shape-symbolic notations and the top nodes correspond to clusters that have been merged in a larger cluster.

The advantages of applying this technique include that it has been proven that the result is unique, and all the membership functions (represented by their shape-symbolic notations) that are contained in a cluster will have a strong relationship (i.e., high degree of similarity) among them.

Although in the remainder of this dissertation only considers the use of the hierarchical algorithm, several well-known clustering algorithms[2] can be used and are subject to further study.

Within this study, the use of the hierarchical algorithm avoids that additional initialization parameters produce different results —e.g., number of clusters, initial cluster centers or seeds. Indeed, the clustering phase focuses on obtaining good partitions representing similar opinions based on the shape-similarity of membership functions.

For a better understanding of how the shape-similarity detection method reduces the complexity of a decision-making problem with regard to the number of opinions given by persons that are involved, a detailed example is given in the next section.

3.4 Illustrative Example

The *shape-similarity detection method* is illustrated in the following example.

Usefulness of a new product. The General Manager of a company likes to encourage the participation of all its staff in order to make a decision. The company has $k = 120$ members who are considered as "experts" by the General Manager when a decision is about the development of a new product based on its "usefulness level" (criterion). Thus, all the experts were asked to supply their opinions over the level of usefulness of a new product.

This study suggests that each expert may express his/her opinion using different domains unified by means of fuzzy sets. In this example these fuzzy sets are limited to the $[0, max]$ domain, and each opinion on the criterion *level of usefulness* is represented by means of a membership function through parameters a, b, c and d.

For testing purposes, the experts' opinions were randomly generated considering different representations of trapezoidal membership functions with $max = 100$. A sample of these values is shown in Table 3.1.

Expert	*4-tuple(a, b, c, d)*
1	(20, 50, 52, 80)
2	(15, 45, 94, 97)
⋮	⋮
k	(20, 53, 53, 85)

Table 3.1: Membership functions represented as 4-tuples (a,b,c,d).

The linguistic term set S^{length} used for this test corresponds to the previously shown on Figure 3.9. Below, the corresponding labels, linguistic terms

[2]Such as K-means [25], ISODATA [26], X-means [27], fuzzy c-means [28] and other adapted algorithms for large scale data [29].

and semantics (represented by triangular membership functions) are included as a reference in Table 3.2.

Label	*Linguistic term*	*Semantic value*
ES	extremely short	(0, 0, 0.17)
VS	very short	(0, 0.17, 0.33)
S	short	(0.17, 0.33, 0.5)
M	medium	(0.33, 0.5, 0.67)
L	long	(0.5, 0.67, 0.83)
VL	very long	(0.67, 0.83, 1)
EL	extremely long	(0.83, 1, 1)

Table 3.2: Linguistic term set and its semantics represented by triangular membership functions.

3.4.1 Phase 1. Obtaining Symbolic Notations

As detailed in Section 3.3.1 the shape-symbolic notation is obtained for each membership function. For instance, one may consider the 4-tuple $(20, 50, 52, 80)$ provided by *expert 1*. Here, the shape-symbolic notation is formed by the shape-string “0+1-0” and the feature-string “VS-S-ES-S-VS”. These strings were obtained by applying algorithms *getShapeString* and *getFeatureString*, respectively.

For illustration purposes, Figure 3.14 shows a mapping of the segments and notation symbols of the membership function μ_1 representing the opinion of *expert 1* during this phase.

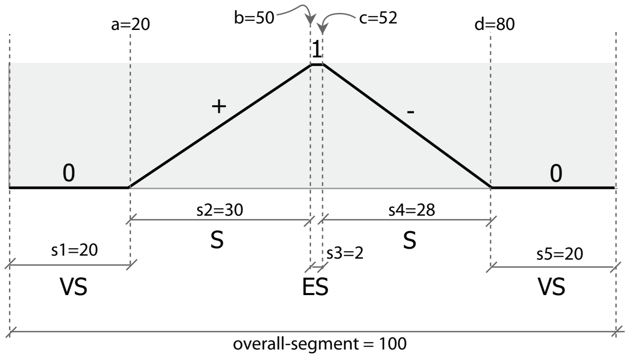


Figure 3.14: Segments of a trapezoidal membership function $\mu_1 = (20, 50, 52, 80)$ to build its shape-symbolic notation.

3.4.1.1 Getting the *shape-string*

Recalling from Section 2.2.1, the 4-tuple $(20, 50, 52, 80)$ corresponds to the trapezoidal membership function

$$\mu_1(x) = \begin{cases} 0 & , \quad x \leq a \\ \frac{x-20}{50-20} & , \quad a < x < b \\ 1 & , \quad b \leq x \leq c \\ \frac{80-x}{80-52} & , \quad c < x < d \\ 0 & , \quad x \geq d \end{cases}$$

The membership function μ_1 and the *sorted-list* P are used as parameters in Algorithm *getShapeString*. The *sorted-list* P is a generalized representation of a membership function that contains its parameters and the domain limits, i.e., for *expert 1* the sorted-list is $P_1 = (0, 20, 50, 52, 80, 100)$.

At the beginning of algorithm *getShapeString* the variables *symbol* and *sStr* are initialized as empty strings (i.e., “”) while variable n with value 6 represents the length of sorted-list P_1. Table 3.3 shows the values of the variables obtained when the aforementioned input parameters were applied.

i	*P*[*i*]	*P*[*i+1*]	*f*(*P*[*i*])	*f*(*P*[*i+1*])	*symbol*	*sStr*
1	0	20	0	0	“0”	“0”
2	20	50	0	1	“+”	“0+”
3	50	52	1	1	“1”	“0+1”
4	52	80	1	0	“-”	“0+1-”
5	80	100	0	0	“0”	“0+1-0”

Table 3.3: Values of variables obtained when applying the *getShapeString* algorithm for input parameters provided by *expert 1*.

At the end of the algorithm the variable *sStr* contains the *shape-string* “0+1-0”.

3.4.1.2 Getting the *feature-string*

For Algorithm *getFeatureString* the aforementioned sorted-list P and the linguistic term set S (Table 3.2) are needed as input parameters.

When the algorithm starts, the *fStr* variable is initialized with an empty string (i.e.,“”), and variable n with value 6 represents the length of sorted-list P_1. The *osl* variable is initialized with zero. After computing the value for the *overall-segment length* according to Equation 3.1, the *osl* variable is obtained as follows:

$$\begin{aligned} osl &= 20 + 30 + 2 + 28 + 20 \\ &= 100. \end{aligned}$$

Afterward, the relative length value of each segment present in μ_1 must be calculated to obtain the linguistic term (*lTerm*) that best represents its relative length. The label of the selected linguistic term is used to build the *featureString*.

Table 3.4 shows the values of the variables obtained when the input parameters corresponding to *expert 1* were applied.

i	*P[i+1]*	*P[i]*	*Segment Length*	*Relative Length*	*lTerm*	*fStr*
1	20	0	20	0.20	Very short	"VS"
2	50	20	30	0.30	Short	"VS-S"
3	52	50	2	0.02	Extremely short	"VS-S-ES'
4	80	52	28	0.28	Short	"VS-S-ES-S"
5	100	80	20	0.20	Very short	"VS-S-ES-S-VS"

Table 3.4: Values of variables obtained when applying the *getFeatureString* algorithm for inputs parameters provided by *expert 1*.

For the sake of clarity, one may consider segment $i = 2$ (i.e., segment $\overline{ab}$) of membership function μ_1 representing the opinion of *expert 1*. The algorithm *getFeatureString* obtains the length value of this segment by the following subtraction:

$$\begin{aligned} segmentLength &= P[i+1] - P[i] \\ &= P[3] - P[2] \\ &= 50 - 20 \\ &= 30. \end{aligned}$$

Next, using Equation 3.2 the algorithm obtains the *relative length* of this segment:

$$r(2) = \frac{L_2(\mu_1)}{osl(\mu_1)} = \frac{30}{100} = 0.30.$$

As mentioned in Section 3.3.1.2, within this dissertation the linguistic term with the highest membership degree for the previously obtained relative length is used. Figure 3.15 shows the linguistic term "Short" labeled as "S" with a higher membership degree.

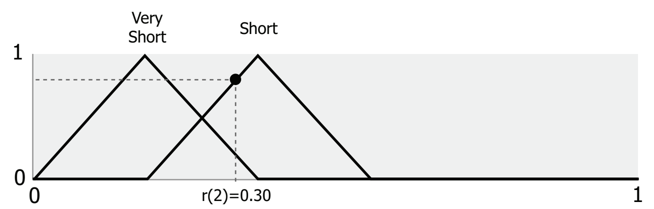


Figure 3.15: Linguistic terms associated to relative length $r(2) = 0.30$.

At the end, variable $fStr$ contains the *feature-string* "VS-S-ES-S-VS".

Consequently, after applying algorithms *getShapeString* and *getFeatureString* the shape-symbolic notations for each 4-tuple provided by the experts are obtained. Table 3.5 shows a sample of the obtained shape-symbolic notations.

Expert	*Shape-symbolic notation*
1	("0+1-0","VS-S-ES-S-VS")
2	("0+1-0","VS-S-M-ES-ES")
⋮	⋮
k	("0+H-0","VS-S-ES-S-VS")

Table 3.5: Sample of the obtained *shape-symbolic notations* during phase 1 of the shape-similarity detection method.

3.4.2 Phase 2. Computing Shape-Similarities

During phase 2, the similarity between all possible pairs of shape-symbolic notations are computed and used to complete the shape-similarity matrix.

In order to illustrate how to calculate the aforementioned similarity, one may consider notations n_8=("1-0","L-VS-VS") and n_{17}=("0+1-0","M-VS-S-ES-ES") representing the opinions of *experts 8* and *17*, respectively.

First, it is necessary to compute the cost of edit operations C_e which includes the insert costs C_i, delete costs C_d and the replace costs C_r as follows:

$$\begin{aligned} C_i &= C_{i_{\langle 0,M\rangle}} + C_{i_{\langle +,VS\rangle}} \\ &= (1+ \mid S^{length} \mid) + (1+ \mid S^{length} \mid) \\ &= 16 \end{aligned}$$

$$C_d = 0$$

$$\begin{aligned} C_r &= C_{r_{\langle 1,L\rangle \frown \langle 1,S\rangle}} + C_{r_{\langle -,VS\rangle \frown \langle -,ES\rangle}} + C_{r_{\langle 0,VS\rangle \frown \langle 0,ES\rangle}} \\ &= 2+1+1 \\ &= 4 \end{aligned}$$

$$\begin{aligned} C_e(n_8, n_{17}) &= C_i + C_d + C_r \\ &= 16+0+4 \\ &= 20. \end{aligned}$$

Second, $e_{max}(n_8, n_{17})$ representing the maximum length between these notations is obtained:

$$\begin{aligned} e_{max}(n_8, n_{17}) &= \max(length(n_8), length(n_{17})) \cdot (1+ \mid S^{length} \mid) \\ &= \max(3,5) \cdot (1+7) \\ &= 40. \end{aligned}$$

Then, the *notation-penalty* that measures the difference between these notations as indicated in Equation 3.4 is calculated:

$$p(n_8, n_{17}) = \frac{C_e(n_8, n_{17})}{e_{max}(n_8, n_{17})} = \frac{20}{40} = 0.5.$$

Finally, the *shape-similarity measure* between the shape-symbolic notations n_8 and n_{17} is obtained using Equation 3.3:

$$S(n_8, n_{17}) = 1 - p(n_8, n_{17}) = 1 - 0.5 = 0.5$$

Analogously, all the shape-similarity measures among the available shape-symbolic notations are computed in order to be used during the next phase.

3.4.3 Phase 3. Grouping By Shape Similarity

During this phase to facilitate the grouping by shape similarity, the previously obtained similarity measures are contained in the *shape-similarity matrix*.

As explained in Section 3.3.3, each shape-symbolic notation initially corresponds to one cluster, and these clusters are merged according to the highest similarity value that is available in the similarity-matrix taking into consideration a specified *threshold*.

Figure 3.16 shows three different groups of trapezoidal membership functions corresponding to the shape-symbolic notations that have been used for this test. These groups represent examples of the resulting clusters in order to reflect the shape-similarity among the membership functions by using a threshold of $\tau = 0.95$.

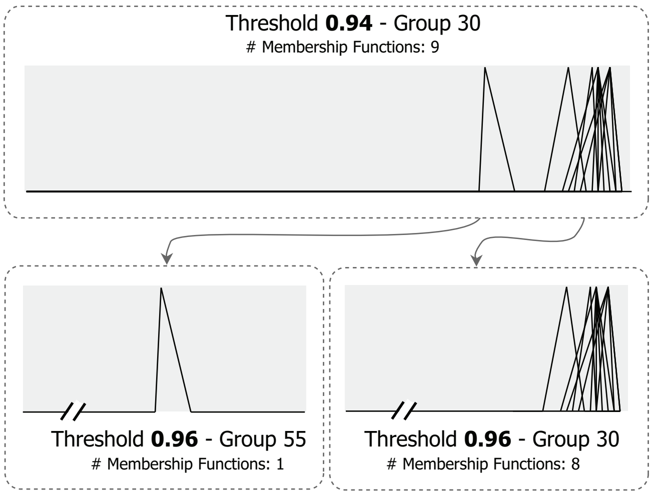


Figure 3.16: Examples of resulting clusters by using a threshold value of 0.95.

Recalling the advantage of the clustering algorithm that is used herein — i.e., obtaining the same result despite of additional initialization parameters—, one may get different cluster configurations by changing the threshold on the resulting dendrogram. Figure 3.17 shows how a *threshold* τ may influence in the choice of a shape-symbolic notation being part or not of a group in a dendrogram's region. For instance, when $\tau = 0.94$ the clusters $\{n_{50}\}$ and $\{n_{30}, n_{31}, n_{32}, n_{33}, n_{42}, n_{40}, n_{46}, n_{55}\}$ are formed. Analogously, when $\tau = 0.96$ all the shape-symbolic notations are contained in the same cluster. In this way, the variation of the threshold allows a decision maker — for instance, the General Manager within this example— to adjust the number of groups or the desired cluster profile —e.g., size and level of similarity.

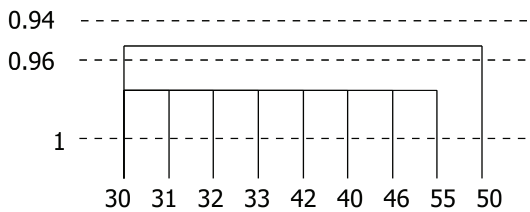


Figure 3.17: A region of the obtained dendrogram.

At the end of this phase, the shape-similarity detection method has reduced the complexity of a decision-making problem from handling k opinions given by experts to m groups of similar opinions where $m << k$.

3.4.4 Obtained Results

To carry out the simulations presented in the example illustrated in this section, a functional prototype was implemented. Herein, the starting point is the presence of 120 opinions regarding the criterion "usefulness of a product". These opinions are considered to be provided by "experts" acting as participants in a decision where all the participants are considered to be equally important. Thus, each opinion was represented by means of a trapezoidal membership function.

Next, the obtained results during the different phases of the *shape-similarity detection method* are presented. Figure 3.18 is used to visualize the three phases of the method: (1) obtaining shape-symbolic notations, (2) computing shape-similarities, and (3) grouping by shape similarity.

In phase 1, 120 shape-symbolic notations corresponding to the 120 membership functions representing expert opinions were obtained. On the left side of Figure 3.18, the mapping between the shape-symbolic notations and the membership functions could be observed. The importance of this phase is to obtain the shape-symbolic notations as an abstraction of the expert opinions represented by trapezoidal membership functions. Since these notations are built using character strings and linguistic terms, it will allow for their comparisons in the next phase. Some results of this phase are shown in Table 3.6.

In phase 2, the shape-similarity measures ranging from 0 to 1 between pairs

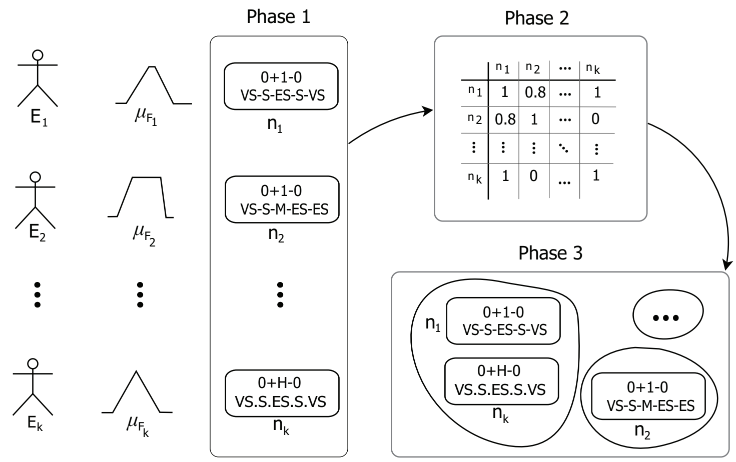


Figure 3.18: Detailed architecture of the shape-similarity detection method.

Expert	*4-tuple(a, b, c, d)*	*Shape-symbolic notation*
2	(15,45,94,97)	(“0+1-0”,“VS-S-M-ES-ES”)
7	(0,0,39,78)	(“1-0”,“S-S-VS”)
8	(0,0,64,86)	(“1-0”,“L-VS-VS”)
17	(49,73,100,100)	(“0+1-0”,“M-VS-S-ES-ES”)
103	(20,48,54,80)	(“0+1-0”,“VS-S-ES-S-VS”)

Table 3.6: Mapping of 4-tuple(a,b,c,d) with their corresponding shape-symbolic notations.

of expert opinions represented by means of shape-symbolic notations were obtained. Considering that the comparison process between notations is based on the number of edit operations, opinions can easily be compared. The following represents a portion of the obtained shape-similarity matrix.

$$
\begin{array}{c c}
 & \begin{array}{ccccc} 2 & 7 & 8 & 17 & 103 \end{array} \\
\begin{array}{c} 2 \\ 7 \\ 8 \\ 17 \\ 103 \end{array} &
\begin{pmatrix}
0.50 & 0.53 & 0.90 & 1.00 & 0.85 \\
1.00 & 0.88 & 0.53 & 0.50 & 0.55 \\
0.88 & 1.00 & 0.50 & 0.53 & 0.48 \\
0.53 & 0.50 & 1.00 & 0.90 & 0.80 \\
0.55 & 0.48 & 0.80 & 0.85 & 1.00
\end{pmatrix}
\end{array}
$$

In phase 3 during the clustering step, different cluster configurations were obtained. To do this, different thresholds ranging from 0.5 to 1.0 were applied. Table 3.7 shows the number of clusters according to the applied threshold.

Threshold τ	Number of groups
0.50	1
0.55	2
0.60	2
0.65	2
0.70	2
0.75	2
0.80	2
0.85	5
0.90	17
0.95	50
1.00	68

Table 3.7: Number of groups according to the applied threshold τ.

3.4.5 Results Interpretation

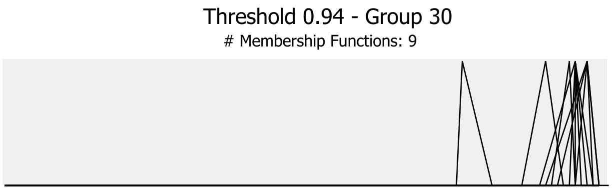
For the analyzed data set with thresholds lower than 0.85, the clustering step was determined by the shape-string. Thus, the existence of two big groups where their elements have a common shape-string representations was observed. The first group correspond to shapes represented by the shape-strings "1-0" and "0+1-0". In the second group, it was possible to notice that membership functions represented by shape-strings "0+1-0" and "0+H-0" converge when the middle segment is "extremely short".

In the case where the thresholds take values higher or equal than 0.85, the clustering step takes into account the feature-string. Thus, it was possible to observe more details regarding the relative lengths of the membership function segments. Therefore, membership functions might still belong to different groups according to the relative length of their segments, despite the fact that these functions sometimes have the same shape-string. Here, it is important to mention that the computation of the relative length of the segments depends on the selected linguistic term set S^{length}. In the example, a set with seven linguistic terms was used. Although the use of a more detailed linguistic term set might detect differences with more accuracy —i.e., applying different levels of granularity— the linguistic term set should not contain too many terms. As it has been mentioned in Section 2.4, the linguistic term set should be defined with caution considering our capacity to process information.

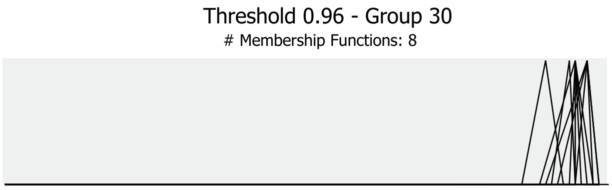
As shown in Table 3.7, a different number of groups could be obtained through different thresholds. Thus, a decision maker may adjust the threshold level according to the number of groups or the cluster profile that he/she would like to consider for further analysis. Herein, the cluster profile refers to its size, which could lead to provide clusters representing a majority or a minority.

To illustrate how the variation of the threshold influences the clustering process, Figure 3.19c depicts two instances of this process with different values. Here, it could be noticed that the shape-symbolic notation n_{55} belongs to group

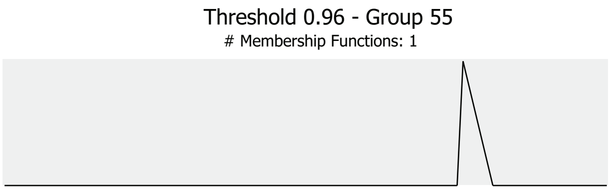
G_{30} with threshold $\tau = 0.94$ while this notation is part of a separate group G_{55} with threshold $\tau = 0.96$.



(a)



(b)



(c)

Figure 3.19: Effect of using different thresholds during the clustering process.

3.5 Conclusions

This chapter has described a novel method, called *shape-similarity detection method*, based on the strategy of grouping a large number of opinions by similarity while answering to the question *how to simplify the complexity of a decision-making problem that involves a large number of persons in the decision?* Herein, it is assumed that the similarity among opinions (i.e., using preference modeling) representing the same concept may be detected by the similarity on the shape characteristics of the membership functions reflecting their preferences. Therefore, a novel representation of the membership functions called *shape-symbolic notation* has been provided.

A *shape-symbolic notation* helps the detection method to reduce the complexity of processing a large number of membership functions (representing opinions) by facilitating their comparisons. The shape-symbolic notation has two components: the *shape-string* and the *feature-string*. The shape-string denotes the shape characteristics of the membership function. The feature-string expresses the relative length of the membership function on the X-axis segments. The steps to properly build these components from a membership function has been provided in this chapter.

To facilitate the comparison among a large number of membership functions, a novel *shape-similarity measure* has been also provided. This measure obtains the similarity degree between two shape-symbolic notations by computing the cost of transforming one shape-symbolic notation into the other for penalty purposes. In addition to the shape-similarity measure, a *shape-similarity matrix* is provided to group similarly shaped membership functions with ease.

By grouping similarly shaped membership functions, the *shape-similarity detection method* reduces the complexity in the presence of a large number of membership functions, therefore this method may be used in a decision-making process that involves a large number of opinions. In other words, this method allows a decision-maker to perceive a few representative groups of opinions instead of a large number of individual ones, in such a way that he/she does not have to deal with all the given opinions, but with a reduced set of grouped opinions. To the best of the author's knowledge this is the first shape based similarity detection method presented in the literature.

How to identify representative opinions from a decision maker's perspective will be described in the next chapter.

References

[1] Ana Tapia-Rosero, Antoon Bronselaer, and Guy De Tré. *A method based on shape-similarity for detecting similar opinions in group decision-making.* Information Sciences, 258:291–311, 2014.

[2] Lofti Zadeh. *Similarity relations and fuzzy orderings+.* Information sciences, 3(2):177–200, 1971.

[3] George J. Klir and Bo Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice Hall, 1995.

[4] Henry Blumberg et al. *Hausdorff's Grundzüge der Mengenlehre.* Bulletin of the American Mathematical Society, 27(3):116–129, 1920.

[5] Richard W. Hamming. *Error detecting and error correcting codes.* Bell System Technical Journal, 29(2):147–160, 1950.

[6] Rolf Niedermeier and Peter Sanders. *On the Manhattan distance between points on space filling mesh indexings.* Universität Karlsruhe, Informatik für Ingenieure und Naturwissenschaftler, 1996.

[7] Per-Erik Danielsson. *Euclidean distance mapping.* Computer Graphics and Image Processing, 14(3):227 – 248, 1980.

[8] F. Chiclana, J.M. Tapia García, M.J. del Moral, and Enrique Herrera-Viedma. *A statistical comparative study of different similarity measures of consensus in group decision making.* Information Sciences, 221:110–123, February 2013.

[9] Dongrui Wu and Jerry Mendel. *A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets.* Information Sciences, 178(2):381–402, January 2008.

[10] Didier Dubois and Henry Prade. *Fundamentals of Fuzzy Sets (The Handbooks of Fuzzy Sets Volume 7).* Springer, 1st edition, 2000.

[11] Amos Tversky. *Features of Similarities.* Psychological Review, 84:327–352, 1977.

[12] Lee R. Dice. *Measures of the Amount of Ecologic Association Between Species.* Ecology, 26(3):297–302, jul 1945.

[13] Paul Jaccard. *Nouvelles recherches sur la distribution florale.* Bulletin de la Sociète Vaudense des Sciences Naturelles, 44:223–270, 1908.

[14] Chengguo Lu, Jibin Lan, and Zhongxing Wang. *Aggregation of Fuzzy Opinions Under Group Decision-Making Based on Similarity and Distance.* Journal of Systems Science and Complexity, 19(1):63–71, March 2006.

[15] Leila Baccour, Adel M. Alimi, and Robert I. John. *Similarity measures for intuitionistic fuzzy sets: State of the art.* Journal of Intelligent and Fuzzy Systems, 24(1):37–49, 2013.

[16] Eulalia Szmidt. *Distances and Similarities in Intuitionistic Fuzzy Sets.* Springer, 2014.

[17] Zeshui Xu and Meimei Xia. *Distance and similarity measures for hesitant fuzzy sets.* Information Sciences, 181(11):2128–2138, June 2011.

[18] Zhi-yong Bai. *Distance similarity measures for interval-valued hesitant fuzzy sets and their application in multicriteria decision making.* Journal of Decision Systems, 22(3):190–201, 2013.

[19] Rami Zwick and Edward Carlstein. *Measures of similarity among fuzzy concepts: A comparative analysis.* International Journal of Approximate, pages 221–242, 1987.

[20] Ana Tapia-Rosero, Antoon Bronselaer, and Guy De Tré. *Similarity of Membership Functions: A Shaped based Approach.* In Proceedings of the 4th International Joint Conference on Computational Intelligence, pages 402–409. SciTePress - Science and Technology Publications, 2012.

[21] Luis Martínez and Francisco Herrera. *An overview on the 2-tuple linguistic model for computing with words in decision making: Extensions, applications and challenges.* Information Sciences, 207:1–18, November 2012.

[22] Rosa M. Rodríguez, Luis Martínez, and Francisco Herrera. *Hesistant Fuzzy Linguistic Term Sets for Decision Making.* IEEE Transactions on Fuzzy Systems, 20(1):109–119, 2012.

[23] Dan Gusfield. *Algorithms on strings, trees, and sequences: computer science and computational biology.* Cambridge University Press, New York, NY, USA, 1997.

[24] Robin Sibson. *SLINK: an optimally efficient algorithm for the single-link cluster method.* The Computer Journal, 16(1):30–34, 1973.

[25] Anil K. Jain. *Data clustering: 50 years beyond K-means.* Pattern Recognition Letters, 31(8):651–666, June 2010.

[26] Geoffrey H. Ball and David J. Hall. *ISODATA, a novel method of data analysis and pattern classification.* Technical report, DTIC Document, 1965.

[27] Dan Pelleg, Andrew Moore, et al. *X-means: Extending k-means with efficient estimation of the number of clusters.* In Proceedings of the seventeenth international conference on machine learning, volume 1, pages 727–734. San Francisco, 2000.

[28] Angélica Urrutia, Hector Valdes, and José Galindo. *Comparison of K-Means and Fuzzy C-Means Data Mining Algorithms for Analysis of Management Information: An Open Source Case.* In Distributed Computing and Artificial Intelligence, pages 187–195. Springer, 2013.

[29] Lawrence O. Hall. *Exploring Big Data with Scalable Soft Clustering.* Synergies of Soft Computing and Statistics for Intelligent Data Analysis, pages 11–15, 2013.

Chapter 4

Identifying and Evaluating Relevant Opinions

Parts of this chapter were published in:

- Tapia-Rosero Ana and De Tré Guy. **A cohesion measure for expert preferences in group decision-making**. In *Modern approaches in fuzzy sets, intuitionistic fuzzy sets, generalized nets and related topics. Volume II: Applications* edited by Krassimir Atanassov, Michal Baczynski, Jozef Drewniak, Janusz Kacprzyk, Maciej Krawczak, Eulalia Szmidt, Maciej Wygralak and Slawomir Zadrozny, 125-142, SRI-PAS, 2014.
- Tapia-Rosero Ana and De Tré Guy. **Evaluating relevant opinions within a large group**. In *Proceedings of the 6th International Joint Conference on Computational Intelligence*, 402-409. Rome, Italy, 2014.

4.1 Introduction

As presented in Chapter 3, the complexity of decision making with regard to the presence of a large number of opinions may be reduced by grouping them by similarity. Once a large group of opinions has been partitioned into subgroups of similar opinions, a decision maker can categorize them according to some of their individual characteristics. An important question surrounding this kind of categorization given by a decision maker is *how to identify and evaluate (groups of) relevant opinions within a large amount of opinions where some of them are more representative than others?* The following example helps to provide some insights into this topic.

The usefulness of a new feature in a product. A company wants to know the "usefulness level" (criterion) of a new pressure sensor (feature) in an electric toothbrush (product) while the product is under design. Here, it is possible to gather this information using social media such as the company's fan page. Considering that these opinions may be given by persons with different backgrounds (i.e., education levels, areas of expertise and personal profiles), it is desired to differentiate the opinions considered to be relevant from a decision maker's perspective.

In the example, the term *relevant* refers to a variety of opinions (expressed as fuzzy preferences) which are significant or important to a particular person acting as a decision maker. Thus, one may consider cases where (i) the opinions of some specific professionals might be more important than the opinions of some regular users, (ii) the opinions given by a majority might be more important than the opinions given by a minority, or (iii) the opinions given by a small group expressing confidence might be more important than a large group expressing doubts.

The first case refers to opinions that are worthy of notice among others for a particular perspective such as the opinions given by experts or persons with a particular background —for instance, the opinion given by a sensor engineer in the electric toothbrush example. Hereafter, for readability purposes one may refer to this kind of opinions as *noticeable*.

The second case corresponds to what is known as a majority approach where most of the opinions are important for a particular perspective —e.g., a decision maker from a financial point of view expects that a large number of customers find the (electric toothbrush) product to be useful. A variant of this case is also possible, that is where the opinions given by a minority group are considered to be more relevant than the opinions given by a majority. For instance, a small group of parents prone to buy a toothbrush with a pressure sensor for their children.

The third case takes into account the level of confidence of each group of opinions in order to be compared to other groups. Although it could be considered that any measure that provides the level of confidence of a group of opinions can be used, this case needs to consider how a large number of opinions (where each opinion is expressed using preference modeling) has been partitioned into subgroups of similar opinions.

Besides the aforementioned cases, other cases may be considered including a combination thereof as depicted by the following analogy: "As well as one bright bulb could light up a room as good as a higher number of less brighter bulbs, one may consider that the opinion of one expert might highlight among others" [1]. Based on this analogy, relevant opinions can be modeled considering that the *representativeness* of a group is obtained by combining the number of opinions and the number of the noticeable ones. In this way, it could be noticed that the representativeness of a group somehow takes into account cases (*i*) and (*ii*). Figure 4.1 depicts a strategy to identify (groups of) relevant opinions according to a decision maker's perspective. Here, the groups considered to be worthy of notice (or relevant) are depicted with a gray background —the more gray the background is the more relevant the group is.

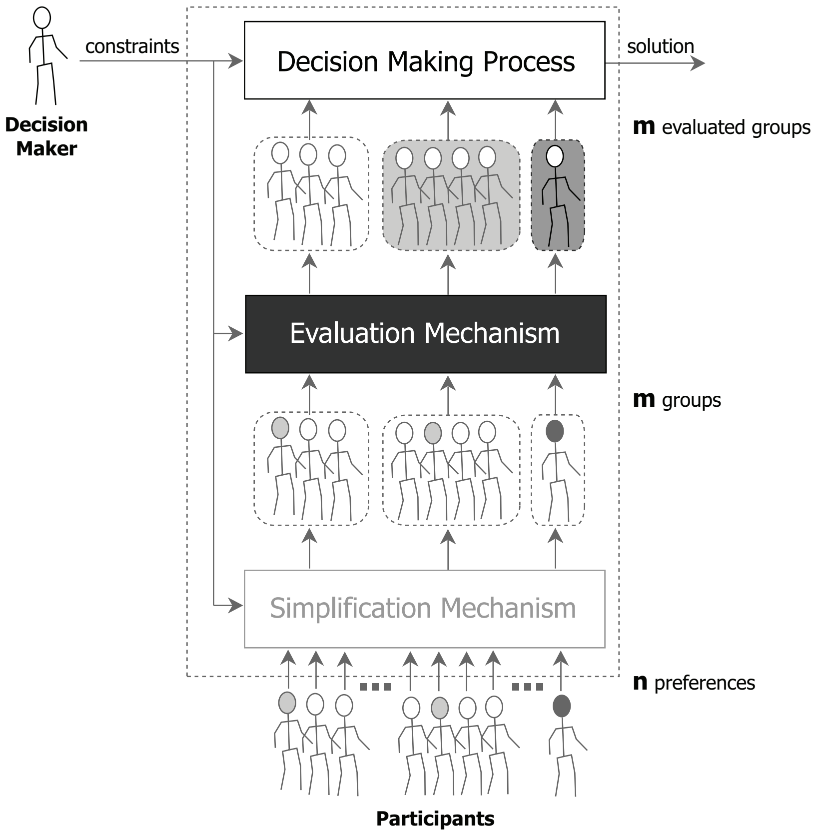


Figure 4.1: A strategy to identify and evaluate groups of relevant opinions (or preferences).

The aim of this chapter is to identify and evaluate (groups of) relevant opinions from a decision maker's perspective. This chapter assumes that the perspective of a decision maker combines the three introductory cases, i.e., a decision maker takes into account the representativeness of a group —cases (i) and (ii)— and in addition its level of confidence —case (iii).

On the one hand, the identification of relevant opinions is based on the available characteristics of the groups —e.g., the size of each group. Since the groups of similar opinions are formed by means of a shape-similarity method, this chapter contributes towards their discrimination by providing two novel approaches to compute a *cohesion measure* which reflects the level of confidence for each group. On the other hand, groups of relevant opinions result from the evaluation of one of the group characteristics (or a combination thereof) according to a decision maker's perspective. Here, the challenge is to reflect someone's perspective based on these characteristics —e.g., the group size, the number of noticeable opinions, and the cohesion measure. Therefore, this chapter also provides a model and the steps to obtain the relevance of each group by combining its characteristics.

The remainder of this chapter is structured as follows. Section 4.2 presents some related work. Section 4.3 describes how to identify groups of relevant opinions. Furthermore, this section provides details on two approaches for computing the *cohesion measure* of a group which are proposed to reflect the level of confidence in groups representing similar opinions —i.e., groups formed by similarly shaped membership functions. Section 4.4 gives a general view of the evaluation of an object on the basis of its characteristics. Additionally, the proposed model and the steps to obtain the relevance of a group are described. Section 4.5 presents the answer to the introductory research question and the contributions of this chapter.

4.2 Related Work

Considering that this chapter is focus on identifying an evaluating relevant opinions, this section aims to provide some related work on this topic.

Among them, the study presented in [2] indicates that "simple averages of judgments of the individual experts is quite effective, and that only a small number of experts must be included to achieve most of the total improvement possible with a much larger set of experts". So, a proper combination of the experts' judgments is quite effective for improving validity.

Considering the size of available web information, the research presented in [3] states that "search engines should be evaluated by their ability to retrieve highly relevant pages rather than all possible relevant pages". Thus, this work explores this aspect by means of a three point relevance scale while selecting the best pages for each analyzed topic. For each topic, multiple assessors (experts) selected the best document and it was found that "assessors frequently disagreed on which document was best". Additionally, it was found that "the relative effectiveness of systems when evaluated by different assessors changed markedly". So, this allows to consider that the identification and evaluation of relevance somehow should take into account the decision maker's point of view.

An interesting work presents a generic and domain independent opinion relevance model for a Social Network user [4]. This proposal estimate the relevance of a single opinion based on different parameters such as the author experience, the experience regarding a particular feature, the word scarcity in an opinion, network distance and similarity. In this case the proposed similarity takes into account the interests in common between an author and an opinion consumer where not necessarily a decision maker's point of view is reflected.

There are some model for evaluating criteria where the term relevance is taken into account. Among them, the model for evaluating a product presented in [5] states that "human evaluators often express their assessment in subjective expressions, in particular, the linguistic terms; for example, to express the relevance of a criterion to a design theme, the terms 'relevant (or important)' and 'very relevant (or very important)' ". However, in this work it is not mentioned how the evaluators perform the evaluation of relevance on criteria. Another work [6], presents a group decision procedure where the experts provide their individual opinions regarding an attribute (or criterion) followed by their aggregation where a subgroup is selected. In this case, the selected individuals correspond "to the members whose collective assessment reach a specified threshold".

The study presented in [7] explores the consensus-relevant information content while provides a framework to support the consensus building. This proposal first determines the preferences of the group (i.e., as a whole) and considers that consensus relevant information is embedded in the preference data. Although one of the steps (during the consensus process) consists in the negotiation using the available consensus relevant information, it is not clear what is considered to be consensus relevant information. In a similar way as other consensus processes, it uses preference relations and a similarity measure to reach a consensus.

In [8], it is considered that the representativeness of the objects differ within a category. Thus, when some objects are more typical than others, it can be considered that these exemplify better their category. The work proposes to characterize data sets by constructing prototypes taking into account two points of views, namely internal and external. In this case, the typicality degrees are computed and the prototype is defined as the aggregation of the most typical data. The typicality degrees of objects are obtained from the internal resemblance and the external dissimilarity among the categories. In this work, the relevance is considered to be an attribute locally defined for each category and this attribute is not defined for the whole data set.

4.3 Identifying Relevant Opinions

In the presence of a large number of opinions that have been grouped by similarity, a decision maker could categorize these groups according to some of their individual characteristics. For instance, the *number of opinions* contained in a group is an attribute that can be used to differentiate (or identify) the groups of similar opinions that represent a majority or a minority.

To identify a group having relevant opinions, one could establish one or more attributes that will be used to compute to what extent the group is relevant. For instance, the aforementioned opinions might have been gathered through several information sources —e.g., fan pages, surveys, or different social network applications— where some profile data of the participants —i.e., the opinion providers— can be available. In this social media context, a decision maker could establish that an important attribute of a group is the average age of its opinion providers. For the purpose of this chapter, it is assumed that a large number of opinions have been partitioned into groups or clusters formed by similarly shaped membership functions as presented in Chapter 3. Hence, these groups are available and the next step is to investigate them.

Among these groups, a decision maker can identify groups of relevant opinions based on the three introductory cases. Therefore, the following attributes have been established to compute the relevance of the groups:

Number of noticeable opinions. This attribute represents the number of opinions that are worthy of notice among others in a group such as the opinions given by experts or persons with a particular background — case (i).

Group size. This attribute indicates the magnitude of a group with regard to the represented opinions –case (ii). In other words, the number of membership functions that are contained in a group.

Cohesion. This attribute is a measure that reflects the level of confidence in a group formed by similarly shaped membership functions —case (iii).

In general, different attributes can be taken into account when identifying groups considered to be relevant from a decision maker's perspective. It could be noticed that some attributes can be directly obtained such as the *group size* and *the number of noticeable opinions*, while the *cohesion* attribute requires some additional computations.

Since the cohesion attribute is related to the strategy that has been used to divide a large number of opinions into groups of similar opinions, two approaches that aim to measure the cohesion of a group are proposed next.

4.3.1 Measuring the Cohesion of a Group

In this dissertation, a *cohesion measure* aims to reflect the level of confidence in a group of opinions formed by means of the shape-similarity detection method. Here, it is important to recall that the shape-similarity detection method provides m groups of similar opinions where their similarity is based on the shape characteristics of the membership functions used to represent opinions on a specific criterion.

The following scenarios illustrate the idea about measuring the cohesion of a group as an indication of its level of confidence:

Several opinions represented by the same membership function.

One could think about a cluster that contains, e.g., one hundred opinions where all the opinion providers *fully agree* with their preferences on a criterion, and hence these opinions are represented by the same membership function. Figure 4.2 illustrates this case where a unique membership function is observable since all the membership functions are drawn one over the other. In this scenario the group has the highest level of togetherness (or agreement on the criterion) among its membership functions, and hence can represent such a group of opinions with the highest confidence level.

Several opinions represented by different membership functions.

One could think about a cluster that contains, e.g., one hundred opinions where some of the opinion providers *partially agree* with their preferences on a criterion, and hence these opinions are represented by different membership functions. This case is illustrated in Figure 4.3. Here, it could be observed

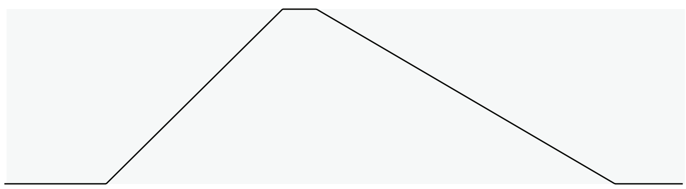


Figure 4.2: A cluster where all the opinion providers *fully agree* with their preferences on a criterion. So, a unique membership function is observable since all the membership functions are drawn one over the other.

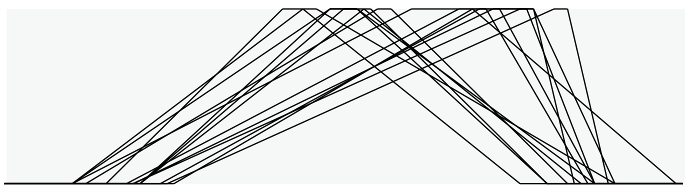


Figure 4.3: A cluster where some of the opinion providers *partially agree* with their preferences on a criterion. So, these opinions are represented by different membership functions.

that the level of togetherness among the membership functions in this group is not the highest.

In a decision-making context, the first scenario represents an ideal but might be unrealistic, while the second scenario might be more expected. This could lead to use a cohesion measure that computes the level of togetherness —or agreement on a criterion— among the membership functions within a group. This should be done in such a way that a higher cohesion might be considered closer to the ideal case.

Before describing the approaches for computing the cohesion of a group, a *cohesion measure* in the context of the shape-similarity detection method (Chapter 3) is formally defined. In [9] a *cohesion measure* is defined as follows.

Definition 4.1 (Shape-Cohesion)
A shape-cohesion *is a togetherness measure among membership functions that belong to the same cluster where the cluster is obtained based on the shape-similarity detection method.*

Two approaches to compute the level of togetherness (or cohesion) among the membership functions contained in a cluster have been studied and proposed in [9], namely the computation by means of an extended shape-symbolic notation and the computation based on a geometric approach.

4.3.1.1 An Extended Shape-Symbolic Notation Approach

In a similar way in which a shape-symbolic notation represents someone's preferences on a criterion (see Chapter 3 Section 3.3.1), an *extended shape-symbolic notation* represents a group of preferences.

To obtain this representation, a group of preferences is considered to be characterized by an upper and a lower bound. Figure 4.4 shows an example of a cluster and its corresponding upper and lower bounds represented by a thick and a dashed line respectively.

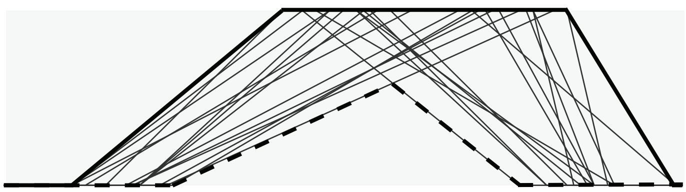


Figure 4.4: An example of a cluster and its corresponding upper and lower bounds.

For the purpose of readability, the upper bound is represented by parameters (a^U, b^U, c^U, d^U) while the lower bound is represented by the parameters (a^L, b^L, c^L, d^L). These parameters are obtained by means of the mathematical operations min and max. Here, one may consider that a cluster is formed by k normalized membership functions where a membership function $\mu_i = (a_i, b_i, c_i, d_i)$ for $i = 1, \ldots, k$. So, the approximations of the upper bound parameters are obtained as follows:

$$a^U = \min(a_1, a_2, \ldots, a_k), \tag{4.1}$$
$$b^U = \min(b_1, b_2, \ldots, b_k), \tag{4.2}$$
$$c^U = \max(c_1, c_2, \ldots, c_k) \text{ and} \tag{4.3}$$
$$d^U = \max(d_1, d_2, \ldots, d_k). \tag{4.4}$$

In a similar way, the approximations of the lower bound parameters are given by

$$a^L = \max(a_1, a_2, \ldots, a_k), \tag{4.5}$$
$$b^L = \max(b_1, b_2, \ldots, b_k), \tag{4.6}$$
$$c^L = \min(c_1, c_2, \ldots, c_k) \text{ and} \tag{4.7}$$
$$d^L = \min(d_1, d_2, \ldots, d_k). \tag{4.8}$$

Considering that each cluster can be enclosed within two boundaries (i.e., defining a range of possible preference values determined by an interval), in this approach a cluster might be seen to some extent as an *interval-valued fuzzy set (IVFS)* [10, 11, 12, 13]. This means that one might think about the *width* of an interval as an indication of the cohesion of a group. Thus, an additional component to represent the *width* is needed in the shape-symbolic notation.

Recalling that the shape-symbolic notation represents each segment of a membership function using two components, namely a *shape-string* and a *feature-string*, the inclusion of a third component, called *width-string*, results in an *extended shape-symbolic notation*. An *extended shape-symbolic notation* is characterized by a triplet:

$$\langle \textit{shape-string, feature-string, width-string} \rangle.$$

The *shape-string* and the *feature-string* are obtained as described in Sections 3.3.1.1 and 3.3.1.2 respectively. To obtain the *width-string* for an 'interval-valued' membership function the following steps are proposed:

- First, a membership function that represents the average between the lower and upper bounds in a cluster is built. This membership function μ^A is denoted by parameters (a^A, b^A, c^A, d^A). For instance, Figure 4.5 shows a membership function μ^A representing the average between the boundaries of a cluster.

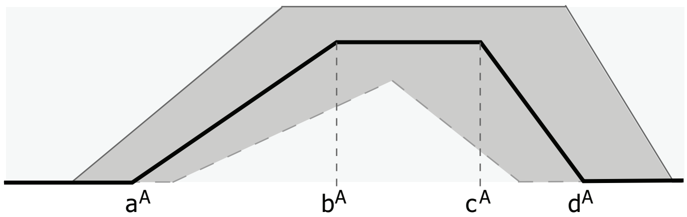


Figure 4.5: A membership function representing the average between the lower and upper bounds in a cluster. This membership function μ^A is denoted by parameters (a^A, b^A, c^A, d^A).

- Second, each segment of the obtained 'average' membership function is approximated to a rectangle. Thus, a rectangle on a segment i has length l_i and width w_i. For instance, in Figure 4.6 these rectangles are drawn around the solid line that represents the average between the lower and upper bounds.

  Here, a rectangle on a segment i is built as follows.

  - To obtain the length l_i on a segment i, the length of the corresponding segment of the average membership function is used. For instance, Figure 4.7 depicts segment $i = 3$ where its length l_3 corresponds to the length between parameters b^A and c^A.

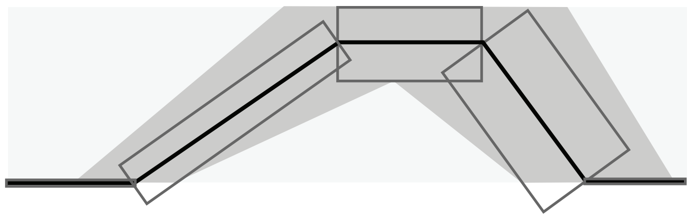


Figure 4.6: A membership function where each of its segments is approximated to a rectangle.

- To obtain the width w_i on a segment i, an approximation of the widest segment is computed. The widest segment corresponds to a segment located between the boundaries while keeping the average membership function in the middle.

  A segment without a slope —i.e., graphically identified as a horizontal segment— is limited by the membership grades of the upper and lower bounds given by μ^U and μ^L respectively. Figure 4.7 depicts width w_3 limited by the membership grades of the upper and lower bounds regarding parameter b, i.e. membership grades given by $\mu^U(b^U)$ and $\mu^L(b^L)$.

  The widest segment on a rectangle with a slope is limited by the parameters of the upper and lower bounds given by (a^U, b^U, c^U, d^U) and (a^L, b^L, c^L, d^L) respectively. In the case of segment $i = 4$ in Figure 4.7, the widest segment is limited by the parameter d^L.

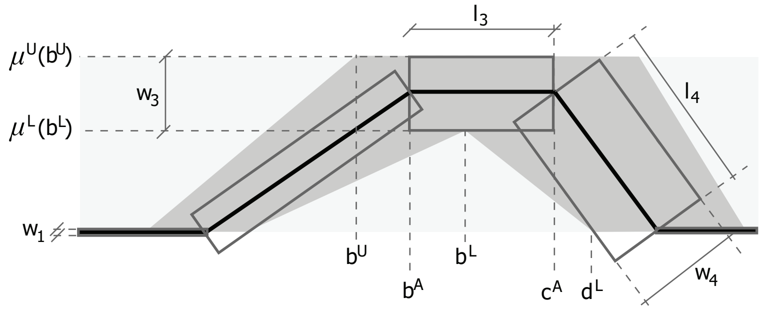


Figure 4.7: An approximation of the widest segments in a membership function representing the average between the lower and upper bounds in a cluster.

- Third, the *width-string* component of the extended shape-symbolic notation associated to the segment i should be selected from the linguistic term-set S^{width}. The linguistic term set S^{width} includes the labels, linguistic terms and semantics indicated in Table 4.1 (cf. the linguistic term set S^{length} in Table 3.2).

Label	*Linguistic term*	*Semantic value*
EN	extremely thin	(0, 0, 0.17)
VN	very thin	(0, 0.17, 0.33)
N	thin	(0.17, 0.33, 0.5)
M	medium	(0.33, 0.5, 0.67)
K	thick	(0.5, 0.67, 0.83)
VK	very thick	(0.67, 0.83, 1)
EK	extremely thick	(0.83, 1, 1)

Table 4.1: Linguistic term set S^{width} and its semantics represented by triangular membership functions.

For instance, an horizontal segment with a range of possible membership grades determined by the $[0, 1]$ interval uses the linguistic term "extremely thick" in its *width-string* component. Analogously, an horizontal segment with a range of possible membership grades determined by the $[0, 0.17]$ interval uses the linguistic term "extremely thin". Figure 4.8 illustrates the use of "extremely thick" and "extremely thin" linguistic terms. Additionally, considering that the *width-string* could be associated with two consecutive linguistic terms, for the sake of simplicity the linguistic term with the highest membership degree is used.

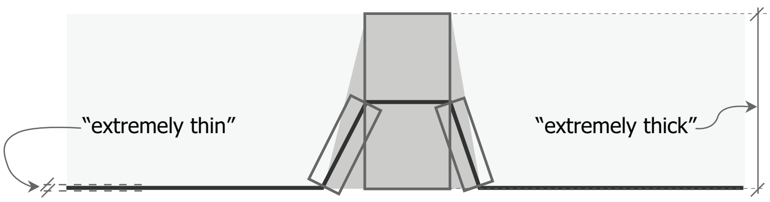


Figure 4.8: An example of "extremely thin" and "extremely thick" segments.

Once all the components of the extended shape-symbolic notation are obtained, the *cohesion* of a group can be computed using

$$cohesion(n_g) = 1 - \sum_{i=1}^{N} l(n_g(i)) \times w(n_g(i)), \tag{4.9}$$

where n_g denotes the extended shape-symbolic notation of cluster g, N is the number of segments that are present in symbolic-notation n_g, $l(n_g(i))$ corresponds to the length component in notation n_g of segment i and $w(n_g(i))$ corresponds to the width component in notation n_g of segment i. As might be noticed, the idea behind Equation 4.9 is to compute a cohesion measure taking into account the numerical values associated to the components of the extended shape-symbolic notation. These components are represented by triangular membership functions herein, so one may consider that these numerical values are the typical values of the corresponding membership functions. For instance, 0.67 is the numerical value associated to a "thick" segment represented by the triangular membership function (0.5, 0.67, 0.83) with typical value $b = c = 0.67$.

Here, one may recall that the extended shape-symbolic notation is characterized by a triplet $\langle$*shape-string, feature-string, width-string*$\rangle$ representing a membership function. The *shape-string* component represents the shape of the membership function (i.e., regarding slopes and preference levels) and it does not include numerical values. Hence, this component is not considered to compute the cohesion measure. The *feature-string* component represents each segment's relative-length by means of linguistic terms in S^{length}. The *width-string* represents each segment's width by means of linguistic terms in S^{width}. Therefore, the *feature-string* and *width-string* components are considered when computing the cohesion measure.

For illustration purposes, one may consider Figure 4.9 which shows cluster g represented by the extended shape-symbolic notation n_g given by:

$$n_g = \langle 0, VS, EN\rangle\langle +, VL, VN\rangle\langle 1, S, N\rangle\langle -, L, M\rangle\langle 0, VS, EN\rangle.$$

Here, the cohesion measure is obtained using Equation 4.9.

$$\begin{aligned}
cohesion(n_g) &= 1 - \textstyle\sum_{i=1}^{5} l(n_g(i)) \times w(n_g(i)) \\
&= 1 - (VS.EN + VL.VN + S.N + L.M + VS.EN) \\
&= 1 - [0.17(0) + 0.83(0.17) + 0.33(0.33) + 0.67(0.5) + 0.17(0)] \\
&= 1 - 0.585 \\
&= 0.415.
\end{aligned}$$

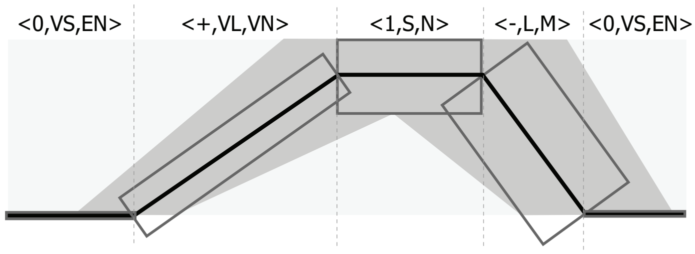


Figure 4.9: Measuring the cohesion of a group using the extended shape-symbolic approach - an example.

4.3.1.2 A Geometric Approach

The starting point for measuring the cohesion of a group by means of a geometric approach is the same as the previous approach, i.e., a group of preferences is considered to be characterized by an upper and a lower bound (Figure 4.4).

In this way, one may consider that these boundaries are represented by two trapeziums as depicted in Figure 4.10. Thus, the upper bound delimits an upper trapezium U referred by the points a^U, b^U, c^U and d^U, where these points are obtained using Equations 4.1, 4.2, 4.3 and 4.4 respectively. Similarly, the lower bound delimits a lower trapezium L referred by the points a^L, b^L, c^L and d^L where these points are obtained using Equations 4.5, 4.6, 4.7 and 4.8 respectively.

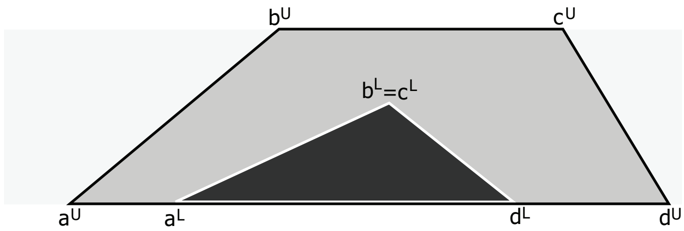


Figure 4.10: Boundaries of a group of preferences characterized by two trapeziums: an upper trapezium U referred by the points a^U, b^U, c^U and d^U, and a lower trapezium L referred by the points a^L, b^L, c^L and d^L.

The idea behind measuring the cohesion in a group using the geometric approach is that the area contained between trapeziums U and L is compared with the area where these polygons are located. Thus, the area of the polygon P formed by the points $a^U, b^U, c^U, d^U, d^L, c^L, b^L, a^L$ and a^U will be an indicator of the cohesion of a group: the larger the area of P, the lower the cohesion; and the smaller the area of P, the higher the cohesion.

Following this idea to compute the cohesion, the next steps are proposed:

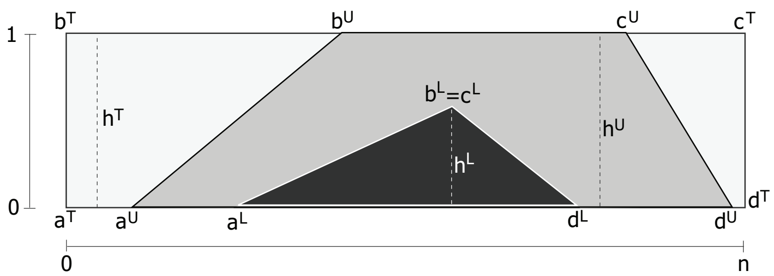


Figure 4.11: Area contained between boundaries (dark gray) compared to the total available area (light gray).

- First, on the basis of the domain of the membership functions that are contained in a cluster, the total area corresponds to the area of the rectangle formed by the points a^T, b^T, c^T and d^T as shown in Figure 4.11.

For instance, in a cluster formed by normalized membership functions given in the domain $[0, n]$ the $area(T) = 1 \times n = n$.

- Second, the area of polygon P is computed by subtracting the area of trapezium L from the area of the trapezium U, i.e. $area(P) = area(U) - area(L)$.
- Third, the cohesion of a cluster g is given by

$$cohesion(g) = 1 - \frac{area(U) - area(L)}{area(T)}. \tag{4.10}$$

For illustration purposes, one may consider Figure 4.11 which shows a cluster g in the domain $[0, 100]$ represented by trapezium U given by $h^T = 1, a^U = 9, b^U = 40, c^U = 82$ and $d^U = 98$, and trapezium L given by $h^L = 0.6, a^L = 25, b^L = c^L = 56$ and $d^L = 76$. Using the steps to compute the cohesion, the following results are obtained:

1. The total area is given by

$$area(T) = 1 \times 100 = 100.$$

2. The area of polygon P is given by $area(P) = area(U) - area(L)$ where

$$area(U) = \frac{(d^U - a^U) + (c^U - b^U)}{2} \times h^T = 65.5$$

and

$$area(L) = \frac{(d^L - a^L) + (c^L - b^L)}{2} \times h^L = 15.3.$$

So, $area(P) = 50.2$.

3. The cohesion measure is obtained using Equation 4.10

$$cohesion(g) = 1 - \frac{area(U) - area(L)}{area(T)} = 0.498.$$

At this point, one may consider that the presented geometrical approach somehow resembles (graphically) to the type-2 fuzzy sets [14]. So, it is worth to mention that what is considered to be the footprint of uncertainty or FOU [15] in type-2 fuzzy sets can be considered to be equivalent to the area between the boundaries as presented herein. In this case, the main difference is that within this dissertation a cluster if formed by a group of similarly shaped membership functions (type-1 fuzzy sets) while the FOU of a single membership function (type-2 fuzzy set) corresponds to a single opinion.

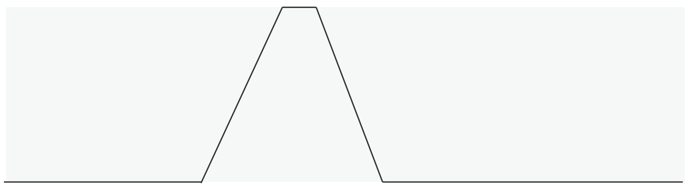


Figure 4.12: Scenario 1. Two opinions represented by the same membership function.

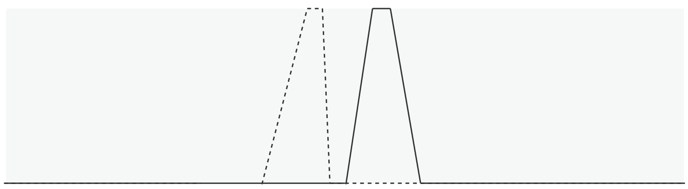


Figure 4.13: Scenario 2. A typical case where two opinions are similar.

4.3.1.3 Comparing the Proposed Approaches

Without loss of generality, to compare the proposed approaches one may consider a cluster formed by two membership functions under the following scenarios:

- *Scenario 1.* This case corresponds to the scenario introduced in this chapter where several opinions are represented by the same membership function. Herein, considering that both membership functions represent a full agreement on a criterion the highest cohesion value is expected. This scenario is depicted in Figure 4.12 and referred to as cluster g_{s1}.
- *Scenario 2.* This case represents a typical case where several opinions mostly agree and hence are represented by similarly shaped membership functions. This scenario is illustrated in Figure 4.13 and referred to as cluster g_{s2}.
- *Scenario 3.* This scenario represents a highly atypical case in which it is assumed that an outlier opinion is included in the cluster. Here, an outlier opinion refers to a dissimilar opinion and hence a very low cohesion measure is expected. This scenario is depicted in Figure 4.14 and referred to as cluster g_{s3}.

For readability purposes, the cohesion measure obtained using the extended shape-symbolic notation approach is referred to as $cohesion_{essn}$ while the cohesion obtained using the geometrical approach is referred to as $cohesion_{geom}$.

Figure 4.14: Scenario 3. A highly atypical case in which is assumed that an outlier opinion is included in the cluster.

Next, each of the aforementioned scenarios are compared by means of the obtained values for $cohesion_{essn}$ and $cohesion_{geom}$.

Scenario 1. The cohesion measures $cohesion_{essn}$ and $cohesion_{geom}$ of cluster $s1$ depicted in Figure 4.12 are obtained as follows.

On the one hand, this cluster is represented by the extended shape-symbolic notation n_{s1} given by:

$$n_{s1} = \langle 0, VL, EN\rangle\langle +, ES, EN\rangle\langle 1, ES, EN\rangle\langle -, ES, EN\rangle\langle 0, VL, EN\rangle.$$

Hence, $cohesion_{essn}$ is computed by:

$$\begin{aligned} cohesion_{essn}(n_{s1}) &= 1 - \textstyle\sum_{i=1}^{5} l(n_{s1}(i)) \times w(n_{s1}(i)) \\ &= 1 - (VL.EN + ES.EN + ES.EN + ES.EN + VL.EN) \\ &= 1 - [1(0) + 0(0) + 0(0) + 0(0) + 1(0)] \\ &= 1. \end{aligned}$$

On the other hand, $cohesion_{geom}$ is obtained using Equation 4.10.

$$cohesion_{geom}(s1) = 1 - \frac{area(U) - area(L)}{area(T)}$$

where $area(U) = area(L)$ and therefore

$$cohesion_{geom}(s1) = 1.$$

In this scenario both approaches provide, as expected, the highest cohesion measure. That is $cohesion_{essn} = cohesion_{geom} = 1$.

Scenario 2. In this scenario $cohesion_{essn}$ is obtained using Figure 4.15, in which it is illustrated that cluster $s2$ is represented by the extended shape-symbolic notation n_{s2}. Accordingly, the resulting notation is

$$n_{s2} = \langle 0, L, EN\rangle\langle +, S, VN\rangle\langle I, S, EK\rangle\langle -, S, VN\rangle\langle 0, L, EN\rangle.$$

Here, $cohesion_{essn}$ is computed by

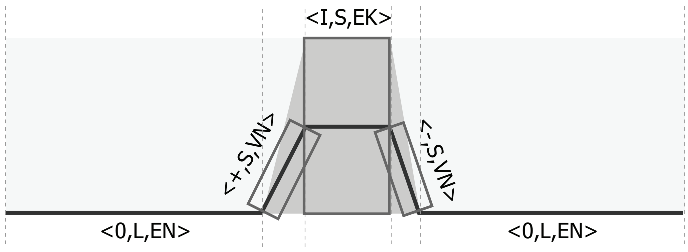


Figure 4.15: Computing $cohesion_{essn}$ in cluster $s2$ represented by the extended shape-symbolic notation n_{s2}.

$$\begin{aligned} cohesion_{essn}(n_{s2}) &= 1 - \textstyle\sum_{i=1}^{5} l(n_{s2}(i)) \times w(n_{s2}(i)) \\ &= 1 - (L.EN + S.VN + S.EK + S.VN + L.EN) \\ &= 1 - [0.67(0) + 0.33(0.17) + 0.33(1) + 0.33(0.17) + 0.67(0)] \\ &= 0.56. \end{aligned}$$

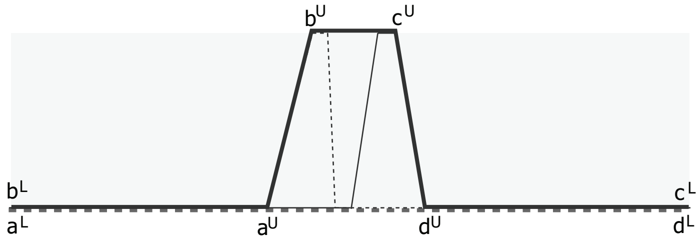


Figure 4.16: Computing $cohesion_{geom}$ in cluster $s2$.

Regarding $cohesion_{geom}$, Figure 4.16 shows cluster $s2$ in the domain $[0, 100]$ represented by trapezium U given by $a^U = 38, b^U = 44, c^U = 57$ and $d^U = 61$, and trapezium L given by $a^L = b^L = 0$ and $c^L = d^L = 100$. In this cluster the height of trapeziums U and L are $h^T = 1$ and $h^L = 0$ respectively. —here, it could also be noticed that trapezium L is a line and hence its area will be zero. Using Equation 4.10 the following results are obtained:

$$cohesion_{geom}(s2) = 1 - \frac{area(U) - area(L)}{area(T)}$$

where $area(L) = 0$, $area(T) = 100$ and $area(U)$ is given by

$$area(U) = \frac{(d^U - a^U) + (c^U - b^U)}{2} \times h^T = \frac{(61 - 38) + (57 - 44)}{2} = 18.$$

Thus,

$$cohesion_{geom}(s2) = 1 - 0.18 = 0.82.$$

Although $cohesion_{essn}$ and $cohesion_{geom}$ differ in the computed values within this scenario, both measures provide an intermediate value as expected. Here, it is worth to mention that an intermediate value can be considered to be more intuitive compared to a zero value, which might have been obtained in the case of using any approach that takes into account the overlapping among the membership functions contained in a cluster as a cohesion measure.

Further work might be performed to suggest which of these measures reflect more properly the perceived cohesion.

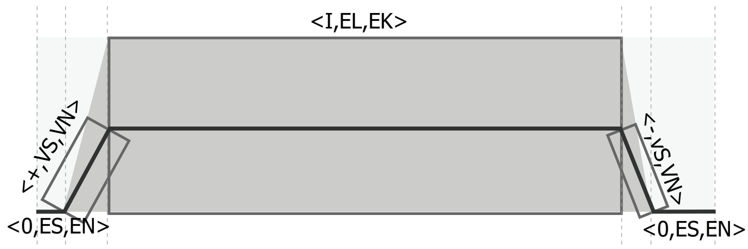


Figure 4.17: Computing $cohesion_{essn}$ in cluster $s3$ represented by the extended shape-symbolic notation n_{s3}.

Scenario 3. The cohesion measure $cohesion_{essn}$ of cluster $s3$, depicted in Figure 4.17, is obtained by means of the extended shape-symbolic notation n_{s3} given by

$$n_{s2} = \langle 0, ES, EN\rangle\langle +, VS, VN\rangle\langle I, EL, EK\rangle\langle -, VS, VN\rangle\langle 0, ES, EN\rangle.$$

Here, $cohesion_{essn}$ is computed as follows:

$$\begin{aligned} cohesion_{essn}(n_{s3}) &= 1 - \textstyle\sum_{i=1}^{5} l(n_{s3}(i)) \times w(n_{s3}(i)) \\ &= 1 - (ES.EN + VS.VN + EL.EK + VS.VN + ES.EN) \\ &= 1 - [0(0) + 0.17(0.17) + 1(1) + 0.17(0.17) + 0(0)] \\ &= 0. \end{aligned}$$

Figure 4.18: Computing $cohesion_{geom}$ in cluster $s3$.

With respect to $cohesion_{geom}$, Figure 4.18 shows cluster $s3$ in the domain $[0, 100]$ represented by trapezium U given by $a^U = 4, b^U = 11, c^U = 86$ and

$d^U = 91$, and trapezium L given by $a^L = b^L = 0$ and $c^L = d^L = 100$. In this cluster the height of trapeziums U and L are $h^T = 1$ and $h^L = 0$ respectively. —here, it could be noticed that trapezium L is a line and hence its area will be zero. In this scenario by means of Equation 4.10 the following results are obtained:

$$cohesion_{geom}(s3) = 1 - \frac{area(U) - area(L)}{area(T)}$$

where $area(L) = 0$, $area(T) = 100$ and $area(U)$ is given by

$$area(U) = \frac{(d^U - a^U) + (c^U - b^U)}{2} \times h^T = \frac{(91 - 4) + (86 - 11)}{2} = 81.$$

Thus,

$$cohesion_{geom}(s2) = 1 - 0.81 = 0.19.$$

In this scenario $cohesion_{essn}$ and $cohesion_{geom}$ differ in the computed values, however both measures provide a very low value as expected.

The aforementioned comparisons between the extended shaped symbolic notation and the geometric approaches allow for noticing that both approaches reflect the expected results. However, considering that the geometric approach provides a more direct computation method, this approach will be used to compute the cohesion hereafter.

Once the cohesion measure of a group has been obtained, this attribute can be used alone or in combination with others to compute an indicator of the relevance of the group as detailed in the next section.

4.4 Evaluating Relevant Opinions

Heretofore different attributes have been identified with the purpose of evaluating groups having relevant opinions. In this section, evaluating groups of relevant opinions consists of (i) computing an indicator of the relevance of each group based of the identified attributes, and (ii) using these indicators to compare the groups. Considering that in these steps the preferences given by a decision maker are taken into account, one can say that the opinions into each group are evaluated according to what is considered to be relevant from a decision maker's perspective.

Accordingly, to compute the *indicator of relevance* of a group, a model and a procedure are described in this section. Before proceeding to present them, a general view of the evaluation of an object on the basis of its attributes (or characteristics) while reflecting a decision maker's perspective is presented. For this purpose, one may consider the following examples where a publisher (acting as a decision maker) has to evaluate several manuscripts (objects) to select the best one for publishing —i.e., the manuscript with the highest relevance.

Evaluation of children's books - Case A. A publisher prefers children's books that have more than five images where eight or more are preferable.

Therefore, the attribute *number of images* is taken into account during the evaluation process.

In this example, the evaluation consists of the computation of $e_{relevance,m}$ as an indicator of the relevance of each manuscript m in a collection of manuscripts, and the use of these indicators to select the best manuscript according to the publisher's preferences. Here, a higher value of $e_{relevance,m}$ denotes a more relevant manuscript and hence a more preferred one. Consequently, the publisher's preferences and a model to compute $e_{relevance,m}$ are needed.

The publisher's preferences can be characterized by the fuzzy set $\mu_{P_{images}}$ shown in Figure 4.19, where $images(x)$ provides the level of preference in the unit interval, i.e. $0 \leq \mu_{P_{images}}(x) \leq 1$, of the variable x representing the number of images in a manuscript.

The model reflecting the publisher's preferences is given by the relevance of a manuscript m regarding the *number of images* in this manuscript, i.e. the model is $e_{relevance,m} = e_{images,m}$ where $e_{images,m} = \mu_{P_{images}}(x)$.

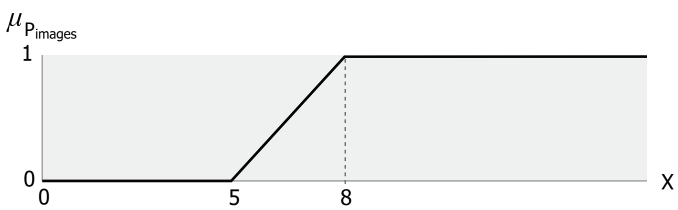


Figure 4.19: Publisher's preferences $\mu_{P_{images}}(x)$ for evaluating children's books according to the attribute *number of images.*

To illustrate the evaluation process, one may consider the case of manuscripts $m1$ and $m2$ with $x = 2$ and $x = 12$ images respectively. Thus, the computation of $e_{relevance,m1}$ and $e_{relevance,m2}$ results from

$$e_{relevance,m1} = e_{images,m1} = \mu_{P_{images}}(x = 2) = 0$$

and

$$e_{relevance,m2} = e_{images,m2} = \mu_{P_{images}}(x = 12) = 1$$

respectively. According to the preferences shown in Figure 4.20, $P_{images}(x = 2) = 0$ and $P_{images}(x = 12) = 1$. Based on the obtained indicators, i.e. $e_{relevance,m1} = 0$ and $e_{relevance,m2} = 1$, manuscript $m2$ is more relevant than manuscript $m1$ from the publisher's perspective.

In a similar way, the evaluation of several manuscripts from the publisher's perspective can be performed.

Evaluation of children's books - Case B. A publisher takes into account two attributes, namely the *number of images* and the *didactical level of the story*, during the evaluation process. The preferences with regard to the *number of images* are kept from the previous example —i.e., manuscripts that

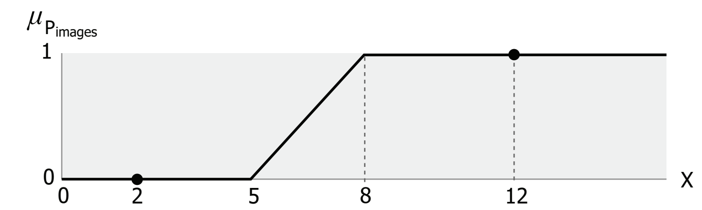


Figure 4.20: Preferences on the number of images of manuscripts $m1$ and $m2$ with $x = 2$ and $x = 12$ images given by $\mu_{P_{images}}(x = 2) = 0$ and $\mu_{P_{images}}(x = 12) = 1$ respectively.

have more than five images where eight or more are preferable. The publisher would like manuscripts with a *'medium' didactical level*, i.e. stories that make a point without being overly didactic. Additionally, the publisher considers that the latter attribute —i.e., didactical level of the story— is two times more important than the number of images.

In this example, the evaluation process is similar to the previous one. However, the publisher's preferences with regard to an additional attribute is needed as well as a different model that allows for the combination of the attributes.

The publisher's preferences with regard to the attribute *didactical level of the story* can be characterized by the fuzzy set $P_{didactic}$ shown in Figure 4.21. Since the preference of this attribute is given by a linguistic term, the semantic value of a linguistic term set[1] that can be used herein is also depicted as a reference through a lighter color.

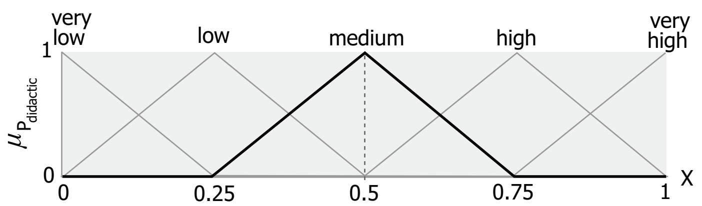


Figure 4.21: Publisher's preferences $\mu_{P_{didactic}}(x)$ for evaluating children's books according to their *didactical level*.

It could be noticed that the preference of a manuscript with regard to the attributes *number of images* and *didactical level of the story* can be individually obtained using the membership functions representing the publisher's preferences on these attributes, i.e. $\mu_{P_{images}}$ and $\mu_{P_{didactic}}$. However, a model to compute the relevance based on both attributes is needed. Thus, the following

[1]A linguistic term set T={L_1='very low', L_2='low', L_3='medium', L_4='high', L_5='very high'}.

model is proposed

$$e_{relevance,m} = w_{images} \times e_{images,m} + w_{didactic} \times e_{didactic,m}$$

where $e_{images,m}$ and $e_{didactic,m}$ are the relevance indicators of the attributes *number of images* and *didactical level of the story* respectively; while w_{images} and $w_{didactic}$ are weights denoting the importance level of each attribute. Moreover, it is considered that each importance weight is represented by a real number between 0 and 1, where the sum of all the attributes' weights summed up 1.

This model uses the well-known weighted sum to combine the individual relevance of the attributes, in which one can obtain the overall relevance $e_{relevance,m}$ for a manuscript m by multiplying the relevance of each attribute by a value representing the *importance weight* assigned to the attribute.

Since the publisher considers that the didactical level of the story is two times more important than the number of images, this suggests that the importance weights of these attributes are $w_{didactic} = \frac{2}{3}$ and $w_{images} = \frac{1}{3}$ respectively. It could be noticed in the aforementioned model that the importance weights, given by the publisher, do not change during the evaluation of different manuscripts.

With the purpose of illustration, one may consider evaluating manuscripts $m1$, $m2$ and $m3$ based on their corresponding attributes number of images and didactical level as detailed in Table 4.2.

	Attributes	
Manuscript Identifier	Number of Images	Didactical Level
m1	4	0.50
m2	12	0.20
m3	12	0.40

Table 4.2: Manuscript examples for evaluating children's books.

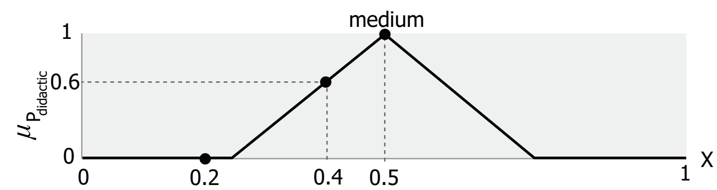


Figure 4.22: Preferences on the didactical level of manuscripts $m1$, $m2$ and m_3 given by $e_{didactic,m1} = P_{didactic}(x = 0.5) = 1$, $e_{didactic,m2} = P_{didactic}(x = 0.2) = 0$, and $e_{didactic,m3} = P_{didactic}(x = 0.4) = 0.6$ respectively.

The individual relevance values of the attribute *didactical level* for manuscripts $m1$, $m2$ and $m3$ are given by their corresponding preferences, i.e.

$e_{didactic,m1} = \mu_{P_{didactic}}(x = 0.5) = 1$, $e_{didactic,m2} = \mu_{P_{didactic}}(x = 0.2) = 0$, and $e_{didactic,m3} = \mu_{P_{didactic}}(x = 0.4) = 0.6$, respectively are shown in Figure 4.22. These values are presented in Table 4.3. It could be noticed that (i) manuscript $m1$ is the least preferred according to its number of images while it is the most preferred according to its didactical level; (ii) manuscripts $m2$ and $m3$ are equally preferred according to their number of images; and (iii) manuscript $m3$ is partially preferred according to its didactical level.

	Individual Evaluation of Attributes	
Manuscript Identifier	e_{images}	$e_{didactic}$
m1	0	1
m2	1	0
m3	1	0.6

Table 4.3: Individual relevance values of the attributes *number of images* and *didactical level* for manuscripts $m1$, $m2$ and $m3$.

By means of the proposed model for this example, the overall relevance values for manuscripts $m1$, $m2$ and $m3$ are $e_{relevance,m1} = 0.67$, $e_{relevance,m2} = 0.33$ and $e_{relevance,m3} = 0.73$ respectively. Based on these values from the publisher's perspective, the best manuscript is $m3$ with the highest overall relevance value, followed with the second highest value for manuscript $m1$, and the third highest value for manuscript m_2.

The examples on the evaluation of children's books provide a general view for evaluating several manuscripts (objects) with one or more attributes from the perspective of a publisher acting as a decision maker. In these examples, the following could be noticed:

- In the presence of one attribute, as shown in *case A*, the overall relevance value of a manuscript is obtained by means of a simple model in which the overall relevance corresponds to the preference on this attribute given by the decision maker.
- In the presence of more attributes, as shown in *case B*, the overall relevance value of a manuscript is obtained by means of a more complex model in which the overall relevance results from the combination of the individual preferences of the attributes and hence an aggregation operator is needed such as the weighted sum.

The ideas behind these examples lead to the design of the model proposed in this dissertation for evaluating groups of relevant opinions.

4.4.1 A Model for Aggregating Preferences on Group Attributes

A graphical representation of the proposed model for aggregating preferences on group attributes, where the groups are formed by similar opinions, is shown in Figure 4.23.

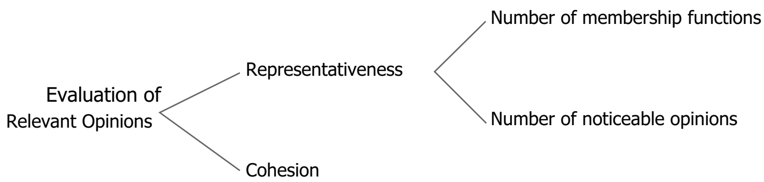


Figure 4.23: A graphical representation of the proposed model for aggregating preferences on group attributes.

This aggregation model is based on three primary attributes, namely the *'number of noticeable opinions'*, *'group size'* and *'cohesion'*, that have been previously established (cf. Section 4.3) to compute the relevance of the groups.

It could be noticed that the aggregation of (the preferences on) the attributes *'number of noticeable opinions'* and *'group size'* results in a new compound attribute called *'representativeness'*. As suggested by its name, this compound attribute denotes the representativeness of a group. Then, by aggregating (the preferences on) the attributes *'representativeness'* and *'cohesion'*, the attribute *'relevance'* is obtained. This attribute is an indicator of the relevance of a group.

Bearing in mind that aggregations are included in the model, any of the operators presented in Section 2.5, i.e. t-norms, t-conorms, ordered weighted averaging (OWA) and the generalized conjunction disjunction (GCD), can be chosen. However, for the purpose of this chapter the generalized conjunction disjunction (GCD) operators [16] have been selected considering that they have different implementations including, among others, a verbalized approach [17] which might be of potential interest in a decision-making context.

Since a decision support technique proposed by Dujmović called *logic scoring of preference (LSP)* [18] allows for the aggregation of multiple criteria based on GCD, the steps implemented in this technique will be used next to perform the aggregation of preferences on group attributes presented in the proposed model.

In the context of the proposed aggregation model, LSP has the following steps: (i) creation of a tree containing the attributes of a group; (ii) characterization of preferences on the attributes given by a decision maker; and (iii) creation of the aggregation structure. These steps are illustrated in Figure 4.24. It could be noticed that, to compute the relevance of a group j, LSP receives as inputs this group and the preferences (or constraints) on the attributes given by a decision maker.

4.4.2 Computing the Relevance of a Group

This section describes the steps to compute the indicator of relevance of a group through an example. The indicator of relevance, or *relevance* for short, is based on the aggregation model proposed in this dissertation and the preferences or constraints given by a decision maker.

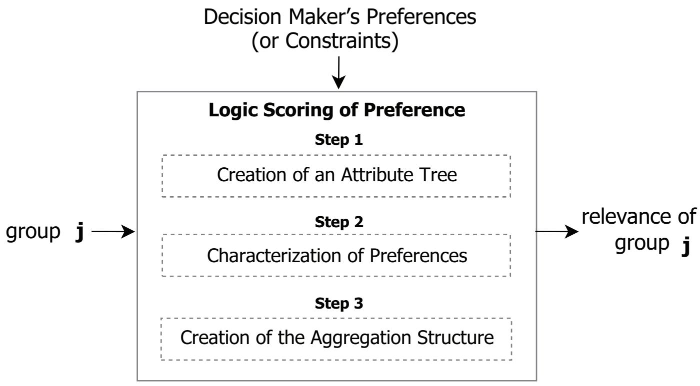


Figure 4.24: LSP steps in the context of a model for aggregating preferences on group attributes from a decision maker's perspective.

For readability, Table 4.4 presents the nomenclature to be used during the computation of the relevance of a group.

Symbol	Designation
i	Identifier of an attribute (or characteristic) in a group
j	Identifier of a group (or cluster) of opinions
$\mu_{P_i}(x)$	Level of preference of variable x regarding attribute i
$v_{i,j}$	Value of attribute i in cluster j
$e_{i,j}$	Relevance of attribute i in cluster j

Table 4.4: Nomenclature used during the computation of the relevance of a group.

At this point, it is assumed that (i) similar opinions have been clustered using a shape based approach, and (ii) the value of the attributes *'cohesion'*, *'number of noticeable opinions'*, and *'size'* (i.e., $v_{cohesion,j}$, $v_{noticeable,j}$ and $v_{size,j}$ respectively) have been computed for each cluster j.

Step 1: Creation of a tree containing the attributes of a group

In this step, the attributes of a group are structured in a tree according to the aggregation model (Figure 4.23). The leaves of the tree represent *primary attributes*, which have been previously measured, are ready to be used, and cannot be further decomposed. In the case of a multilevel attribute tree, each intermediate node represents a *compound attribute* that corresponds to the aggregation of one or more attributes —which could be primary or compound.

In the proposed model, the primary attributes *cohesion*, *number of noticeable opinions* and *group size* are the leaves of the tree, a compound attribute called *representativeness* is an intermediate node, and a compound attribute *relevance* is the root —which will be the indicator of the relevance of a group.

As noticed, the aggregation model has been formalized as an attribute tree within this step. Although the attribute tree of the proposed model is rather simple, an attribute tree allows for the creation of more complex structures where a large number of attributes can be present. Hence the presence of multiple levels and intermediate nodes can be handled using this technique.

Step 2: Characterization of preferences on the primary attributes given by a decision maker

In this step, the preferences on the primary attributes given by a decision maker are characterized through membership functions to reflect his/her perspective regarding the computation of the relevance of a group. Thus, a membership function $\mu_{P_i(x)}$ determines the level of preference of variable x reflecting the preferred values of attribute i.

For instance, Figure 4.25 shows membership function $\mu_{P_{\text{cohesion}}}$ representing that a decision maker accepts clusters with *cohesion* ≥ 0.4 but he/she prefers *cohesion* ≥ 0.6. Furthermore, the decision maker considers that lower values, i.e. *cohesion* < 0.4, are not acceptable.

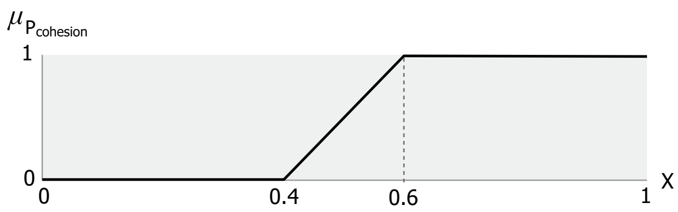


Figure 4.25: Decision maker's preference regarding attribute *'cohesion'* given by $\mu_{P_{\text{cohesion}}(x)}$.

In a similar way, as depicted in Figure 4.26, membership functions $\mu_{P_{\text{noticeable}}}$ and $\mu_{P_{\text{size}}}$ determine the preferences regarding attributes *'number of noticeable opinions'* and *'group size'* respectively.

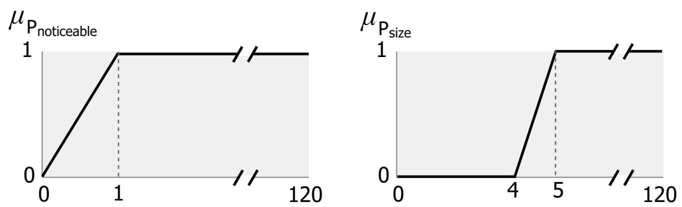


Figure 4.26: Decision maker's preference regarding attributes *'number of noticeable opinions'* and *'group size'* given by $\mu_{P_{\text{noticeable}}(x)}$ and $\mu_{P_{\text{size}}(x)}$ respectively.

Once the preferences of the primary attributes have been characterized, it is possible to compute the *relevance* of these attributes in a cluster. Thus, considering that $e_{i,j}$ corresponds to the *relevance* of attribute i in cluster j and $v_{i,j}$ represents the value of this attribute, therefore $e_{i,j} = \mu_{P_i}(v_{i,j})$.

For the purpose of illustration, one may consider a cluster j and the value of its primary attributes, namely *'number of noticeable opinions'*, *'group size'* and *'cohesion'*, given by $v_{\text{size,j}} = 22$, $v_{\text{noticeable,j}} = 0$ and $v_{\text{cohesion,j}} = 0.4455$ respectively. In this case, the relevance of each primary attribute in cluster j will be obtained using its corresponding membership function to reflect the preferences given by the decision maker. Thus, the following values are obtained:

$$e_{\text{noticeable,j}} = \mu_{P_{\text{noticeable}}}(v_{\text{noticeable,j}}) = \mu_{P_{\text{noticeable}}}(0) = 0,$$

and

$$e_{\text{size,j}} = \mu_{P_{\text{size}}}(v_{\text{size,j}}) = \mu_{P_{\text{size}}}(22) = 1.$$

Additionally, to compute $e_{\text{cohesion,j}}$ corresponding to the relevance given by the attribute *'cohesion'* one could use a linear approximation using

$$\mu_{P_{\text{cohesion}}}(x) = \frac{x - 0.4}{0.6 - 0.4},$$

as follows:

$$e_{\text{cohesion,j}} = \mu_{P_{\text{cohesion}}}(0.4455) = \frac{0.4455 - 0.4}{0.6 - 0.4} = 0.2275.$$

This strategy could be used, as illustrated in Figure 4.27, considering that the preferences are expressed by means of a trapezoidal membership function. Thus, after defining certain dividing points between segments in a function — i.e., parameters a, b, c and d— one could use linear interpolation between them, and as it has been mentioned in [18] "this approach yields a good combination of simplicity and accuracy".

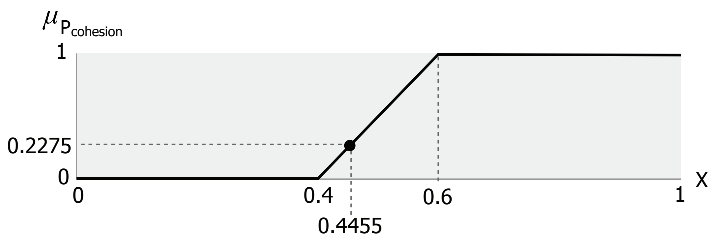


Figure 4.27: Computing $e_{\text{cohesion,j}}$ corresponding to the relevance given by the attribute *'cohesion'* when the value of this attribute is $v_{\text{cohesion,j}} = 0.4455$.

In the case that the *relevance* of a group is based on a single attribute — similar to the '*evaluation of children's books - case A*' example— within this step the *relevance* of all the clusters regarding that attribute can be obtained as described in Section 4.4. Therefore, each cluster j will have an indicator of relevance $e_{i,j}$ for attribute i, where the cluster with the most relevant opinions regarding a single attribute would be the cluster with the highest $e_{i,j}$ value, followed with the second highest value, and so on.

In the case that the *relevance* of a group is based on two or more attributes —similar to the '*evaluation of children's books - case B*' example— within this step the *relevance* for each primary attribute of a group can be obtained using their corresponding functions reflecting the decision maker's preferences. Hence, the *relevance* of all the clusters regarding all their primary attributes can be obtained in the same way. In other words, each cluster j will have an indicator of relevance $e_{i,j}$ for each primary attribute i.

To obtain the cluster with the most relevant opinions regarding two or more attributes, an additional step that performs the aggregation on the previously obtained relevance indicators (i.e., regarding the group attributes) is needed.

Step 3: Creation of the aggregation structure

In this step an *aggregation structure* is created considering that the proposed model needs to aggregate the relevance indicators given by three attributes of a group j, i.e. $e_{\text{size,j}}, e_{\text{noticeable,j}}, e_{\text{cohesion,j}}$. Thus, to obtain an overall relevance indicator $e_{\text{relevance,j}}$, the aggregation structure is established in such a way that, while the aggregation operators satisfy the decision maker's preferences, the structure is consistent with the attribute tree established in Step 1.

For instance, to obtain the compound attribute *'representativeness'* of cluster j it is necessary to take into account its components (i.e. attributes *'number of noticeable opinions'* and *'group size'*) and the level of simultaneity or replace-ability among them. Herein, the *level of simultaneity* refers to the fact that both components need to be satisfied —e.g., a decision maker may prefer a cluster with both a minimum number of noticeable opinions *and* a minimum group size. In contrast, the *level of replace-ability* means that a high representativeness of a cluster can be given by either the 'number of noticeable opinions' *or* 'group size' because one component can replace the other —e.g., a cluster can be considered representative if it has either a minimum number of noticeable opinions *or* a minimum group size.

For the purpose of illustration, Figure 4.28 shows the aforementioned *representativeness* given by $e_{\text{representativeness,j}}$. This indicator is the result of aggregating its components, given by $e_{\text{noticeable,j}}$ and $e_{\text{size,j}}$. Additionally, the *overall relevance* for cluster j given by $e_{\text{relevance,j}}$ can be obtained by aggregating its components $e_{\text{cohesion,j}}$ and $e_{\text{representativeness,j}}$. In this figure, the level of simultaneity or replace-ability has been represented with a blank circle considering that the selection of the corresponding aggregation operators need to be further detailed.

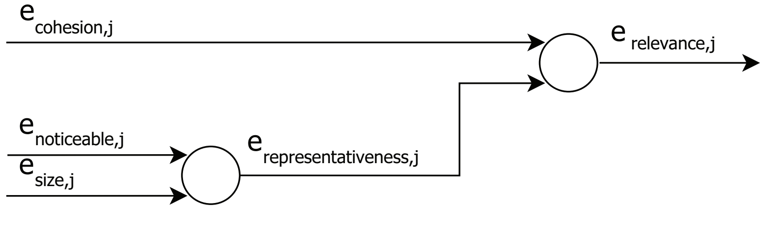


Figure 4.28: Example of the structure to aggregate the relevance indicators $e_{\text{cohesion,j}}, e_{\text{noticeable,j}}, e_{\text{size,j}}$ of group j.

Therefore, to complete that structure, the aggregation operators that satisfy the decision maker's preferences must be selected. In this chapter, the aggregation operators are selected using the GCD verbalized approach (Section 2.5.3.1) which allows the decision maker to use the overall importance scale shown in Table 4.5.

Level	**Overall importance**
16	Highest
15	*Slightly below highest*
14	Very high
13	*Slightly above high*
12	High
11	*Slightly below high*
10	Medium-high
9	*Slightly above medium*
8	Medium
...	...
0	Lowest

Table 4.5: Overall importance scale for GCD verbalized approach.

For example, a decision maker considering that the *'cohesion'* and the *'representativeness'* in a cluster should be to some extent simultaneously satisfied —this implies that there is a level of simultaneity between these attributes. In this case, the decision maker may express the importance of each attribute using Table 4.5.

For the purpose of illustration, the importance of the *cohesion* has been established with a 'High' level (i.e., $l_{\text{cohesion}} = 12$), and the *representativeness* has been established with a 'Medium high' level (i.e., $l_{\text{representativeness}} = 10$). Recalling from Section 2.5.3.1, the simultaneity level α between these attributes is computed using Equation 2.3 as follows:

$$\alpha = \frac{l_{\text{cohesion}} + l_{\text{representativeness}}}{n \cdot L} = \frac{12 + 10}{2 \cdot 16} = 0.6875,$$

where $n = 2$ corresponds to the number of attributes that are involved in the computation, and $L = 16$ corresponds to the maximum level in the overall importance scale (Table 4.5). Here, the obtained value, i.e. $\alpha = 0.6875$, expresses the conjunction degree between the attributes *'cohesion'* and *'representativeness'*.

It could be noticed that, even though the cohesion and the representativeness of a cluster should be simultaneously satisfied, the computed α value is not too high compared to the maximum (i.e., $\alpha = 1$). Thus, the partial conjunction operator represented by the symbol *C-+* is used. This operator is obtained by scanning the column *Andness* (α) in Table 2.1 to look for the computed α-value as shown in Figure 4.29.

Symbol	**Orness**(ω)	**Andness**(α)	**Exponent r**
D	1	0	$+\infty$
D++	0.9375	0.0625	20.63
D+	0.8750	0.1250	9.521
D+-	0.8125	0.1875	5.802
DA	0.7500	0.2500	3.929
D-+	0.6875	0.3125	2.792
D-	0.6250	0.3750	2.018
D–	0.5625	0.4375	1.449
A	0.5	0.5	1
C–	0.4375	0.5625	0.619
C-	0.3750	0.6250	0.261
$\boxed{\text{C-+}}$	0.3125	$\boxed{0.6875}$	-0.148
CA	0.2500	0.7500	-0.72
C+-	0.1875	0.8125	-1.655
C+	0.1250	0.8750	-3.510
C++	0.0625	0.9375	-9.06
C	0	1	$-\infty$

Figure 4.29: Finding the aggregation operator by scanning column *Andness* to look for the computed value $\alpha = 0.6875$.

In a similar way the aggregation operator for the compound attribute *'representativeness'* is obtained on the assumption that the decision maker considers that the representativeness can be given by either the *number of noticeable opinions* or the *group size*. Consequently, the level of replace-ability needs to be computed. On the assumption that the importance levels of the *number of noticeable opinions* is 'High' (i.e., $l_{\text{noticeable}} = 14$), and the importance of the *group size* is 'Very high' (i.e., $l_{\text{size}} = 12$), the level of replace-ability ω between these attributes is computed using Equation 2.4 as follows:

$$\omega = \frac{l_{\text{noticeable}} + l_{\text{size}}}{n \cdot L} = \frac{14 + 12}{2 \cdot 16} = 0.8125,$$

where $n = 2$ corresponds to the number of attributes that are involved in the computation, and $L = 16$ corresponds to the maximum level in the overall importance scale (Table 2.1). Here, the obtained value, i.e. $\omega = 0.8125$, expresses the disjunction degree between the attributes *'number of noticeable opinions'* and *'group size'*. In this case, the aggregation operator *D+-* is used. This operator can be obtained by means of Table 2.1, where one can look for the $\omega = 0.8125$ value in the corresponding *Orness* (ω) column.

Next, the decision maker has to select the weight of each attribute. For instance, if the *cohesion* is two times more important than the *representativeness* then the weights for these attributes are $w_{\text{cohesion}} = 0.67$ and $w_{\text{representativeness}} = 0.33$ respectively. In a similar way, if the components for the *representativeness* are equally important then their weights are 0.5 (i.e., $w_{\text{noticeable}} = w_{\text{size}} = 0.5$). Hence, the aggregation structure to evaluate relevant opinions within a large group, including the weight of each attribute, is shown in Figure 4.30.

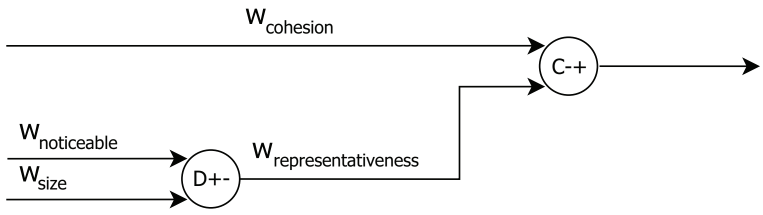


Figure 4.30: Aggregation structure based on the decision maker's perspective where: (i) attributes *'noticeable'* and *'size'* are equally important, i.e. $w_{\text{noticeable}} = w_{\text{size}} = 0.5$; (ii) attribute *'cohesion'* is two times more important than the attribute *'representativeness'*, i.e., $w_{\text{cohesion}} = 0.67$ and $w_{\text{representativeness}} = 0.33$.

Using the previously obtained aggregation structure it is possible to obtain a single value representing the overall relevance indicator for a cluster (Equation 2.2). For instance, the overall relevance for cluster j is obtained as follows. First, the *representativeness* given the selected aggregator *D+-* is computed.

$$
\begin{aligned}
e_{\text{representativeness,j}} &= (w_{\text{noticeable}} \cdot e^r_{\text{noticeable,j}} + w_{\text{size}} \cdot e^r_{\text{size,j}})^{\frac{1}{r}} \\
&= (0.5 \cdot 0^{5.802} + 0.5 \cdot 1^{5.802})^{\frac{1}{5.802}} \\
&= 0.887393.
\end{aligned}
$$

Then, in a similar way, using aggregator *C-+* the overall relevance of cluster j is computed as follows:

$$
\begin{aligned}
e_{\text{relevance,j}} &= (w_{\text{cohesion}} \cdot e^r_{\text{cohesion,j}} + w_{\text{representativeness,j}} \cdot e^r_{\text{representativeness}})^{\frac{1}{r}} \\
&= (0.67 \cdot 0.2275^{-0.148} + 0.3 \cdot 0.88739^{-0.148})^{\frac{1}{-0.148}} \\
&= 0.34610.
\end{aligned}
$$

Thus, the evaluation of cluster j is given by the previously obtained value, i.e. $e_{\text{relevance,j}} = 0.34610$.

It could be noticed that the aggregation structure makes it possible to change with ease different parameters, given by the decision maker, for a better representation of his/her perspective. For instance, if the decision maker would have changed the given weights in the aggregation structure, i.e. $w_{\text{cohesion}} = 0.33$ and $w_{\text{representativeness}} = 0.67$, the overall evaluation value would have been $e_{\text{relevance,j}} = 0.5490246$.

Following the aforementioned steps, an indicator of relevance $e_{\text{relevance,j}}$ are obtained for each cluster j. Therefore, the cluster with the most relevant opinions regarding the aggregation model is the cluster with the highest $e_{\text{relevance,j}}$ value, followed with the second highest value, and so on.

4.5 Conclusions

This chapter proposed to identify and evaluate (groups of) relevant opinions from a decision maker's perspective where some of the opinions are considered to be more representative than others.

On the assumption that groups of similar opinions have been formed, the identification of relevant opinions is based on the following characteristics of the groups: (i) the *number of noticeable opinions* representing the opinions that are worthy of notice in a group, (ii) the *group size* indicating the number of represented opinions in a group, and (iii) the *cohesion* which is a measure of confidence in groups formed by similarly shaped membership functions.

Since the groups of similar opinions are formed by means of the shape-similarity detection method, this chapter contributes towards their discrimination by providing two novel approaches to compute a *cohesion measure*, namely the computation using an extended shape-symbolic notation and the computation using a geometric approach. Both approaches reflect the expected results under their comparison using three different scenarios.

The evaluation of relevant opinions result from obtaining relevance indicators based on one of the group characteristics (or a combination thereof) according to a decision maker's perspective. Considering that it is a challenge to reflect someone's perspective, this chapter has provided a novel aggregation model which is used for computing the relevance indicators of a group. Although to compute the relevance of a group the GCD aggregation operators have been used, other operators could be considered and are subject of further study.

References

[1] Ana Tapia-Rosero and Guy De Tré. *Evaluating relevant opinions within a large group.* In 6th International Conference on Fuzzy Computation Theory and Applications, pages 76–86. SciTePress, 2014.

[2] Robert H. Ashton. *Combining the judgments of experts: How many and which ones?* Organizational Behavior and Human Decision Processes, 38(3):405–414, 1986.

[3] Ellen M. Voorhees. *Evaluation by Highly Relevant Document.* Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, (Vlc):74–82, 2001.

[4] Allan Diego Silva Lima and Jaime Simao Sichman. *SORM: A Social Opinion Relevance Model.* 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), pages 78–85, aug 2014.

[5] Jie Lu, Jun Ma, Guangquan Zhang, Yijun Zhu, Xianyi Zeng, and Ludovic Koehl. *Theme-based comprehensive evaluation in new product development using fuzzy hierarchical criteria group decision-making method.* IEEE Transactions on Industrial Electronics, 58(6):2236–2246, 2011.

[6] Miguel Ángel Ballester and José Luis García-Lapresta. *Sequential Consensus for Selecting Qualified Individuals of a Group.* International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 16(supp02):57–68, 2008.

[7] Noel Bryson. *Group decision-making and the analytic hierarchy process: exploring the consensus-relevant information content.* Computers & Operations Research, 23(1):27–35, 1996.

[8] Marie-Jeanne Lesot, Maria Rifqi, and Bernadette Bouchon-Meunier. *Fuzzy prototypes : from a cognitive view to a machine learning principle.* Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, pages 431–452, 2008.

[9] Ana Tapia-Rosero and Guy De Tré. *A Cohesion Measure for Expert Preferences in Group Decision-Making.* In Krassimir Atanassov, Władysław Homenda, Olgierd Hryniewicz, Janusz Kacprzyk, Maciej Krawczak, Zbigniew Nahorski, Eulalia Szmidt, and Sławomir Zadrożny, editors, New Trends in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics, volume II: Applications, pages 125–142. Systems Research Institute Polish Academy of Sciences, 2013.

[10] Ivor Grattan-Guinness. *Fuzzy Membership Mapped onto Intervals and Many-Valued Quantities.* Mathematical Logic Quarterly, 22(1):149–160, 1975.

[11] Von K.-U. Jahn. *Intervall-wertige Mengen.* Mathematische Nachrichten, 68(1):115–132, 1975.

[12] Roland Sambuc. *Fonctions and floues: application a l'aide au diagnostic en pathologie thyroidienne.* Faculté de Médecine de Marseille, 1975.

[13] Lotfi Zadeh. *The concept of a linguistic variable and its application to approximate reasoning-I.* Information Sciences, 8(3):199–249, January 1975.

[14] Jerry M. Mendel and RI Bob John. *Type-2 fuzzy sets made simple.* IEEE Transactions on fuzzy systems, 10(2):117–127, 2002.

[15] Jerry Mendel and Dongrui Wu. *Perceptual Computing: Aiding People in Making Subjective Judgments.* Wiley-IEEE Press, 2010.

[16] Jozo Dujmović and Henrik Larsen. *Generalized conjunction/disjunction.* International Journal of Approximate Reasoning, 46(3):423–446, December 2007.

[17] Jozo Dujmović. *Andness and orness as a mean of overall importance.* In Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on, pages 1–6. IEEE, 2012.

[18] Jozo Dujmović, Guy De Tré, and Nico Van De Weghe. *LSP suitability maps.* Soft Computing, 14(5):421–434, June 2010.

Chapter 5

Fusion of preferences from different perspectives

Parts of this chapter were published in:

- Tapia-Rosero Ana, Bronselaer Antoon, De Mol Robin and De Tré Guy. **Fusion of preferences from different perspectives in a decision-making context**. *Information Fusion*, 29 (2016): 120-131.

5.1 Introduction

These days, solving a decision-making problem might include preferences from a high number of persons (e.g., potential customers of a product) and managerial constraints (or preferences) given by managers with regard to different aspects (i.e., economical, technical, environmental) over multiple criteria (e.g., features of a product). An interesting question within this regard is *how to handle the complexity of a decision-making problem that involves heterogeneous preferences in the presence of different perspectives?* The next example might provide some better insights on this topic.

A decision regarding the features of a new product. A company has to decide the proper combination of features (criteria) —like capacity, weight and color— during the design of a brand-new model of 'hand luggage' (product). Herein, a proper combination of features should take into account the preferences given by its potential customers, in addition to the managerial constraints (or preferences) given by the company's managers.

Bearing in mind a feature such as 'capacity', it is possible that *a group of users* who mostly travel for short periods using only hand luggage (i.e., business trips), might prefer a "medium capacity". Meanwhile, a *marketing manager* might prefer a "small capacity" with the purpose of promoting that the product will be neither measured at the aircraft entrance, nor placed in the aircraft hold for a specific flyer profile such as the economy class. If the decision involves the whole organization, other perspectives (from other managers) might be present in addition to the one given by the marketing manager. Moreover, depending on the organizational structure of the company, the importance of those perspectives might differ —for instance, a horizontal company treats the opinions of its managers as equally important.

Considering that at the present time several entities —such as governments and businesses— are increasingly interested in gathering opinions from different sources (e.g., fan pages, surveys, polls and social network applications), these opinions might differ somehow in their *relevance* from a decision maker's point of view. As presented in Chapter 4, to reach a decision, it is possible to select the opinions (i.e., preferences) considered to be relevant from *only one* particular perspective. However, in the presence of different perspectives (e.g., from different managers), it is necessary to properly combine the opinions considered to be relevant from each perspective. Therefore, this chapter presents a novel decision-making model that performs a fusion of preferences by means of a *decision-making fusion tree*. Here, the preferences are considered to be given by a large number of participants, as well as by multiple decision makers —where the preferences given by decision makers could also be seen as constraints. Moreover, the model is suitable for different (multilevel) organizational structures. Figure 5.1 depicts a strategy to perform the combination of heterogeneous preferences (provided by a large number of participants) in the presence of different perspectives (provided by multiple decision makers). In this figure, the decision maker* represents a decision maker who provides his/her preferences or constraints regarding the overall decision making.

This dissertation contributes to the study of decision problems that involve different domains of knowledge, i.e. different perspectives given by persons with different areas of expertise, where it is possible to delegate the sub-tasks of decision making. For example, a decision-making problem in a multinational corporation with operations in more than one country, where the headquarters take into account the opinions given by the regional (and sub-regional) organizational units and their respective customers. In this example, each regional (or sub-regional) manager may include in his/her perspective the regional constraints (e.g., cultural, environmental, financial, among others) related to his/her competence area.

Consequently, the proposed decision-making model uses a concept that allows for the fusion of preferences from each perspective. This concept is presented as a *decision-making unit (DMU)* where a single decision maker is able to fuse his/her preferences (or constraints) with the ones received as inputs. The latter might come from a large group of persons involved in a decision, whose preferences may have been gathered from different sources, or other *DMUs* in a hierarchical structure. In this way, the proposed decision-making model allows for splitting a complex problem in manageable sub-problems in a hierarchical fashion.

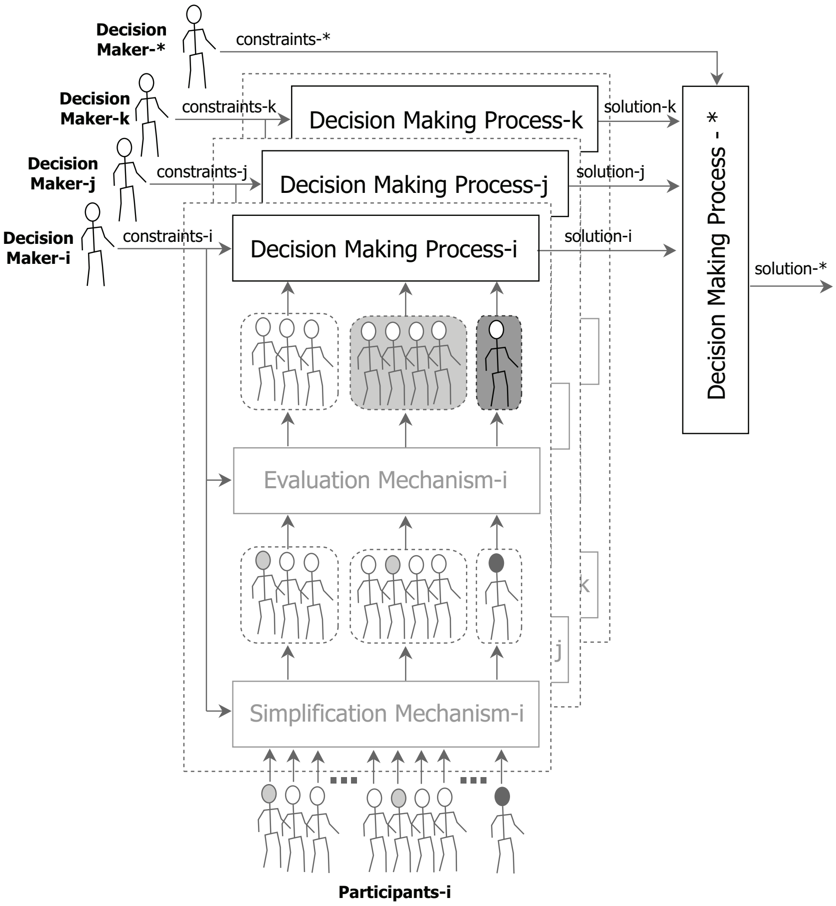


Figure 5.1: A strategy to perform the combination of heterogeneous preferences (provided by a large number of participants) in the presence of different perspectives (provided by multiple decision makers).

An advantage of the proposed decision-making model is that it handles a large group of opinions obtained from different sources where all the decision makers contribute to some extent to the final decision. Furthermore, it evaluates different perspectives separately, and permits a decision maker to obtain a solution that best suits his/her constraints (or preferences) and the preferences given by a large group of persons that are involved. Here, it is a challenge to fuse preferences from different perspectives, while reflecting each decision maker's point of view according to his/her knowledge, experience or area of expertise.

The remainder of this chapter is structured as follows. Next section presents a brief review of related work. Section 5.3 provides details regarding a model that allows for the fusion of preferences from different perspectives in a decision making process which represents the main contribution of this chapter. Additionally, this chapter introduces the concept of a *decision-making unit (DMU)* which is used as the primary component of this model. Section 5.4 illustrates the applicability of the proposed model through an example in the context of a new product design. And, Section 5.5 presents the answer to the introductory question and a discussion on the novel contributions of this chapter.

5.2 Related Work

Multicriteria decision-making (MCDM) problems, i.e. decision-making problems in the presence of multiple criteria, might involve several decision makers in the process. In this case, it is necessary to combine the opinions of all the members in a group on all alternatives under all criteria. This case is known as *multicriteria group decision making (MCGDM).*

In MCGDM is common the use of weights representing levels of importance among criteria as well as among decision makers —for instance, indicating to which degree a decision maker may be more important than another. The solution of MCDM or MCGDM problems consists in obtaining a ranking of alternatives. To find a solution, several methods can be found in the literature including the *analytic hierarchy process (AHP)* [1] and the *technique for the order preference by similarity to ideal solution (TOPSIS)* [2].

Considering the hierarchical representation of the AHP method, this method is described as follows.

Analytic Hierarchy Process (AHP) method. To model a problem, AHP uses a hierarchic structure for its representation and uses pairwise comparisons to establish relations within the structure [3]. The output of this method is a ranking of k alternatives according to their levels of preference expressed in the $[0, 1]$ interval. These levels of preference are computed by means of running pairwise comparisons of alternatives. To do this, an expert (or decision maker) uses a finite scale of values—usually from 1 to 7. While value 1 expresses that two alternatives are equally preferred, higher values express that one alternative is more preferred than the other. These values correspond to entries of a reciprocal matrix $R = [r_{ij}]$ for $i, j = 1, \ldots, k$. The value on the diagonal of the matrix r_{ii} is 1 and the values located symmetrically with respect to the diagonal satisfy the reciprocal condition $r_{ij} = 1/r_{ji}$. In this method, it is possible that the matrix presents some inconsistencies regarding the transitivity property. However AHP allows for the quantification and monitoring of these inconsistencies. In the presence of several experts (or decision makers) the method slightly differs in the computation of a performance index. Here, the performance index corresponds to the sum of the inconsistency indexes for all reciprocal matrices $R[1], R[2], \ldots, R[c]$ on the assumption of “c” reciprocal matrices.

A fuzzy analytic hierarchy process method is presented in [4], which evaluates subjective expert judgments (or opinions) made by a technical committee. This work presents the selection process of government-sponsored R&D projects where different experts provide different points of view, knowledge and assumptions regarding the selection process.

The AHP based on the idea of granularity of information is detailed in [5], where "flexibility is brought into the AHP structures by allowing the reciprocal matrices to be *granular* rather than numeric". In this case, by granular reciprocal matrices refers to matrices whose entries can be information granules, i.e., intervals, fuzzy sets, rough sets, probability density functions, among others.

An interesting approach consists in an extension of the AHP method presented in [6], which uses the OWA operators, where these two approaches operate a different levels. The OWA operators are used to aggregate the children criteria, instead of the weighted average, within the evaluation process in the AHP. In this way, the AHP includes a quantifier guided aggregation. Additionally, a review of the main developments in the analytic hierarchy process is presented in [7], while [8] presents a comparison of some aggregation techniques using group AHP based on three new developed measures.

Another work presents a fuzzy hierarchical criteria group decision-making method [9]. The model handles multiple sources of information, such as persons and sensors. These information sources provide subjective and objective measurements regarding the criteria, which is hierarchically organized. This method uses a linguistic scale to express the levels of importance among the criteria.

Some related works on merging expert opinions include a study [10] where a framework for combining expert opinions (or judgments) is presented with regard to the questions 'how many?' and 'which ones?'. A stochastic process for merging opinions with increasing information is presented in [11]. Additionally, an interesting comparison of algorithmic methods using synthetic data is provided in [12] where approaches that take into account the importance of the experts (i.e., expert weights) have a better performance.

How experts select an alternative from a group of previously generated alternatives based on preferences expressed by means of numerical and linguistic information is described in [13]. In this case a fusion operator for numerical and linguistic information is also presented.

The interest reader may refer to [14] regarding several models for fuzzy multicriteria decision making in addition to the previously presented. However, to the best of the author's knowledge these works do not include handling a large number of opinions provided by regular participants in the decision-making process. Therefore, the next section presents a novel decision-making model that allows the fusion of preferences from different perspectives where the preferences provided by a large number of regular participants are also taken into account.

5.3 A Decision-Making Model for Fusion of Preferences from Different Perspectives

In a decision-making context, in which there are several persons with divergent interests involved, the final decision over the best alternative(s) might be more representative. Therefore, this section presents a novel decision-making model whereby the preferences from a large group of people (like potential customers) and the preferences (or constraints) from different decision makers are fused. In this way, different perspectives (i.e., economical, technical, environmental, etc.) given by different participants in a decision, over multiple criteria, contribute to some extent to the final decision.

In a business environment, an organizational structure generally reflects how the members of a company participate in a decision-making process. Thus, a company with a horizontal organizational structure promotes the participation of all its members and eliminates layers of intermediate management —i.e., between employees and the general manager. This kind of organization usually considers all the opinions of its members to be equally important. Meanwhile, a company with a hierarchical structure includes several managers at different levels and commonly considers that each intermediate manager (expressing his/her perspective) has a different importance degree. For instance, preferences given by the General Manager are considered to be more important than those given by other (intermediate) managers. Besides these two organizational structures, some organizations might conceive a different structure for a specific product, which can also be efficiently handled by the proposed model.

In order to fuse preferences from different perspectives, the proposed decision-making model has to deal with the following problems: (i) the representation of preferences that might be gathered through different sources; (ii) the fusion of preferences reflecting different managerial constraints; and, (iii) the propagation of preferences throughout an organizational structure until the level in which a decision is made. In the next subsections, these problems are more detailed and handled.

5.3.1 Representation of Preferences

In this dissertation, it is assumed that different persons (participating in a decision-making process) may express their preferences using different domains (i.e., numerical, interval-valued or linguistic), due to the presence of different levels of knowledge, areas of expertise and personal profiles. In this approach, these preferences are unified by means of fuzzy sets in a common domain (Section 2.3.1) represented by trapezoidal (or triangular) membership functions.

Considering that a large number of participants are involved, it is necessary to simplify their preferences and hence reducing the complexity of the problem (i.e., from a large number of preferences to a lower number of preference trends). Therefore, preferences representing similar opinions over a criterion are grouped as described in Chapter 3 (i.e., by means of a shape-similarity detection method [15]). In the presence of multiple criteria, the clustering is performed over each criterion separately. Herein, each cluster represents a group of similar preferences over a *criterion* denoted by $G_{criterion,j}$ where j corresponds to the cluster identifier.

Additionally, when the aforementioned preferences are gathered through several information sources, where some profile data of the participants could also be available, it is possible to attach additional data to the clusters. For example, the number of membership functions in a cluster or the proportion gathered by each source, might be considered for representativeness purposes. In a similar way, the area of expertise given by the occupation of the participants might be considered for categorizing some preferences as more relevant than others. In this dissertation, more specifically in Chapter 4 during the identification of relevant opinions, these additional data were referred to as attributes. Hereafter, for readability purposes, it is considered that each cluster $G_{criterion,j}$ has l attributes and one may refer to the attribute i in cluster j as $a_{i,j}$ (Figure 5.2).

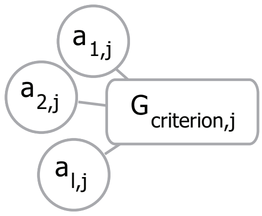


Figure 5.2: Representation of a cluster representing preferences over a *criterion* denoted by $G_{criterion,j}$ where j corresponds to the cluster identifier. This cluster has l attributes, where each attribute is denoted by $a_{i,j}$ for $i = 1, \ldots, l$.

5.3.2 Fusion of Preferences

The aim of this chapter is to study the fusion of preferences given by persons having different perspectives, among them, a decision maker. To achieve this, a proper representation is needed. Therefore, the concept of a *decision-making unit (DMU)* is introduced and becomes the primary component of the proposed decision-making model.

Definition 5.1 (Decision-Making Unit (DMU))
A Decision-Making Unit (DMU) *is an entity where the preferences received from* clients *are fused with the preferences (or constraints) provided by a single decision maker producing an output. Here, a* client *is someone who expresses his/her preferences without deciding what the result is.*

The following describes in more detail the inputs, the processing and the output of a DMU (Figure 5.3).

5.3.2.1 Inputs of a DMU

The inputs of a DMU are preferences given by both the clients and the decision maker, where the preferences of the latter might also be seen as managerial constraints.

For instance, one may consider a company with a multilevel organizational structure as the one depicted in Figure 5.4. Within this organization, a final decision will be made by the General Manager (Level 2) taking into account

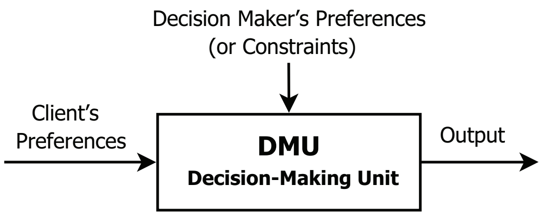


Figure 5.3: Diagram of a decision-making unit (DMU).

the preferences given by its potential customers, and the preferences given by the intermediate managers (Level 1). In this example, there are three DMUs with their corresponding *clients* and *decision makers* as follows. At level 1, there are two DMUs where the potential customers are the clients for both DMUs. Additionally, the Design Manager is the decision maker of one DMU, and the Marketing Manager is the decision maker of the other. The third DMU is located at level 2, where the intermediate managers are the clients and the General Manager is the decision maker. Here, the dual behavior of each intermediate manager could be noticed, namely, as decision maker at level 1 and as client at level 2.

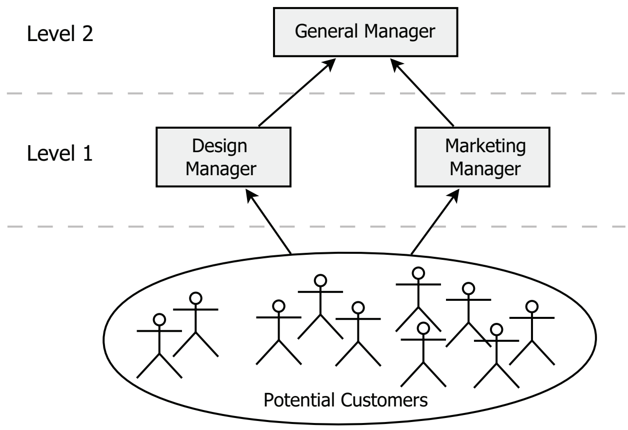


Figure 5.4: Example of a multilevel organizational structure.

Because the proposed decision-making model copes with preferences from a large number of persons (e.g., potential customers or followers of a company on social media), the client's preferences (input of a DMU) might be represented by clusters (i.e., where each cluster represents a group of similar preferences) over a criterion (e.g., usefulness of a product). Therefore, in a multicriteria problem, the client's preferences of a DMU might include several clusters (i.e., several groups of preferences) for t criteria (e.g., color, size, weight, durability of a product), where each cluster G_j has l attributes (see Figure 5.2) —Recalling from Section 4.3, the attributes of a cluster such as number of noticeable opinions, group size and cohesion allow for identifying relevant opinions. Herein, it is important to mention that the number of clusters for each criterion may vary, but the number of attributes is the same for each cluster in a criterion specification. Figure 5.5 depicts each criterion specification as a rectangle —i.e., t rectangles for t criteria— by means of several clusters with l attributes. Each rectangle has a label C_i where i corresponds to the criterion identifier. Additionally, considering that the preferences given by a decision maker reflect an individual point of view over criteria, the decision maker's preferences are expressed by means of a fuzzy set for each criterion —i.e., there are t fuzzy sets for t criteria.

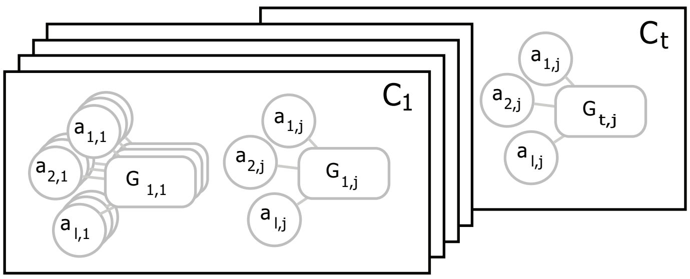


Figure 5.5: Client's preferences over t criteria expressed by clusters of similar preferences, where each cluster $G_{criterion,j}$ has l attributes and each criterion i is depicted as a rectangle C_i for $i = 1, \ldots, t$.

5.3.2.2 Processing in a DMU

The processing in a DMU consists in the fusion of all the preferences received as inputs (i.e., the preferences given by clients and the preferences given by the decision maker) and aims to produce a more informative output reflecting the decision maker's point of view. Thus, the processing in a DMU includes two steps: (i) the *pre-processing of inputs* where the client's preferences are prepared for their further processing, and (ii) the *fusion of preferences* where the preferences are scored from the decision maker's point of view, and then fused with the inputs of the DMU according to the decision maker's requirements. Although different multicriteria methods[1] could be used in this step, for the purpose of this chapter the logic scoring of preference (LSP) method is used.

Figure 5.6 depicts the two steps during the processing in a DMU, which are described below:

1. The *Pre-Processing of Inputs* allows a DMU to adapt the client's preferences for further processing. That is, either combining the information

[1] For a comparative study on different multicriteria methods including the simple additive scoring (SAS), multiattribute value technique (MAVT), multiattribute utility technique (MAUT), analytic hierarchy process (AHP), ordered weighted average (OWA), outranking methods (ELECTRE and PROMETHEE), and logic scoring of preferences (LSP), the interested reader may refer to [16].

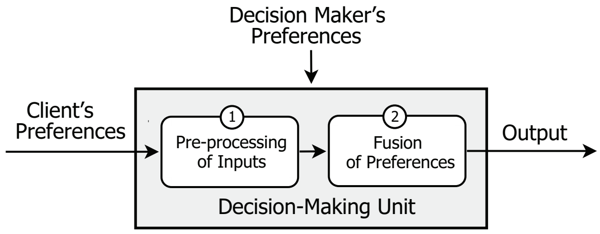


Figure 5.6: Steps during the processing of a DMU.

gathered from multiple sources or adapting the inputs to be used within the fusion step. The former includes to unify the information in a common domain, while the latter includes to prepare a decision-making model that contains the decision maker's preferences to reflect his/her point of view over the inputs. For example, when the fusion step uses the LSP method (Section 4.4.1) the pre-processing of inputs includes the *creation of an attribute tree* when two or more attributes of the client's preferences are taken into account.

2. The *Fusion of Preferences* uses a method to reflect the decision maker's point of view over the client's preferences (e.g., an indicator of relevance or a score), and fuses the new (or updated) information with the input preferences of the DMU. Herein, the new information corresponds to an indicator of relevance (or a score value) in the unit interval $[0, 1]$ expressing to what extent the decision maker prefers the input given by the client. It is important to notice that in the presence of several clusters —representing the client's preferences— all of them should be scored.

For the sake of readability, in this dissertation the fusion of a new attribute will be denoted as a^s_m where the s stands for *score attribute* from the point of view of decision maker m (when a single membership function expressing the client's preference is available). In the case that the client's preferences are clusters, the new attribute will be denoted $a^s_{m,j}$ where j refers to the cluster identifier (in addition to the previously explained identifiers m and s). During the fusion step, it is also possible that the processing of a DMU results in updating previously received attribute values. Hence, in the case of a modified (or updated) attribute in a cluster, it will be denoted by means of an apostrophe. For example, when the attribute $a_{i,j}$ is updated it will be denoted as $a'_{i,j}$.

During the processing in a DMU one could obtain as a result several clusters of preferences, with their corresponding scores, when the client's preferences were given by several clusters as well. Therefore, the pre-processing step might include a strategy for selecting the best group(s) of preferences. For example, in a hierarchical structure, when a DMU becomes a client of another DMU it is possible (i) to select a single cluster, i.e. the one with the highest score; (ii) in a similar way, to select the top K clusters; or (iii) to select all the clusters with a score above a threshold defined by the decision maker —notice that the preferred threshold values might be expressed by means of a fuzzy set as well.

Bearing in mind that the proposed model has been conceived to support diverse methods, the details regarding the processing in a DMU are available through the example given in Section 5.4 when using the logic scoring of preferences (LSP) particularly.

5.3.2.3 Output of a DMU

The output of a DMU consists of clusters of preferences indicating to which level these are considered to be preferred from the decision maker's point of view or considered to satisfy the decision maker's constraints. Such a level is considered to be an additional attribute of each cluster named *score*. The *score* attribute is a value in the unit interval, i.e. $[0, 1]$. Figure 5.7 depicts a DMU diagram with its detailed output which includes the score attribute $a^s_{m,j}$ for each cluster G_j over t criteria. Herein, there are t rectangles for t criteria labeled C_i where i corresponds to the criterion identifier.

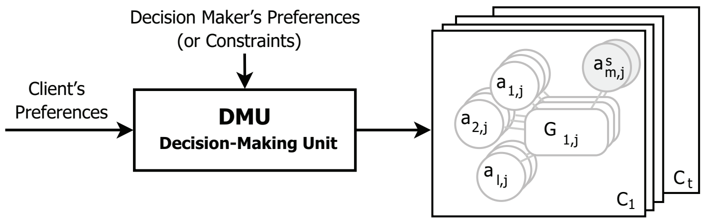


Figure 5.7: Detailed output of a decision-making unit (DMU) where the *scores* given by decision maker m are depicted by means of gray circles.

In a hierarchical structure, the output of a DMU might provide inputs to other DMUs. In this way, the decision makers of a certain level prepare the preferences for the next (superior) level. Thus, decision makers of higher levels, i.e. managers in a higher organizational level, could make a more informed decision.

5.3.3 Propagation of Preferences

The propagation of preferences throughout an organizational structure consists in enriching information, at the level, where a final decision will be made. Hence, the proposed decision-making model propagates preferences (that have been fused) from different perspectives in a multilevel organizational structure based on the DMU concept. In this chapter, to represent an organization, a tree structure T is used where its nodes are DMUs. Hereafter, one can refer to this structure as a *decision-making fusion tree*. Additionally, the DMU class depicted in Figure 5.8 is used for propagating preferences throughout the structure T.

The DMU class includes the following properties:

- *DMmodel* which corresponds to the model that a decision maker will use to reflect his/her point of view;

- *parent* and *children* which allows a DMU to step through the structure T;
- *inputs* corresponding to the client's preferences and the decision maker's preferences, and
- *fusionist* corresponding to the DMU's fusion method —e.g., the LSP aggregation method that is used in this chapter.

Additionally, the DMU class includes two methods: *PreProcess* and *DoFusion* that are used in the DMU processing.

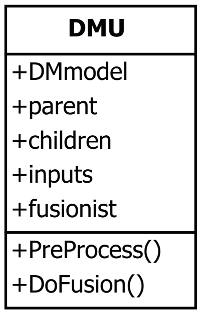


Figure 5.8: DMU class.

Algorithm 5.1 shows the steps that propagate preferences in a multilevel organization given a *dmu* as an instance of the DMU class. The algorithm traverses all the children of a *dmu* (line 1) before their respective parents —i.e., post-order traversal. Line 2 makes a recursive call to ensure that the *dmu's children* are also traversed. Line 3 calls Algorithm 5.2 in order to process the *dmu*. Finally, when the condition in line 4 is satisfied —i.e., when the *dmu* has a parent— the *result* of processing the *dmu* is assigned to the parent's inputs. It is important to notice that, in this way, the *dmu* processes its *inputs* given by its children, and these processes the *inputs* from their children in the same way.

Algorithm 5.1 Propagation of Preferences

Require: dmu.parent
Ensure: dmu.result $\neq NULL$
1: **for all** *child* in *dmu.children* **do**
2:     $child.Algorithm5.1(dmu)$
3: $dmu.Algorithm5.2$ //DMU processing
4: **if** $dmu.parent \neq NULL$
5:     $dmu.parent.inputs \leftarrow dmu.result$

Algorithm 5.2 performs the processing in a decision-making unit as described in Section 5.3.2.2, namely the *pre-processing of inputs* and the *fusion of preferences*. Line 1 preprocesses the inputs according to its attributes $DMmodel$ and $fusionist$. Once the inputs were preprocessed, these will be available at *dmu.inputs*. Line 2 performs the fusion method based on the *DMmodel*. Thus, *DoFusion* obtains a score indicating to what extent the *dmu.inputs* are considered to be preferred from the decision maker's point of view (or considered that satisfied the decision maker's constraints), and fuses this score with *dmu.result*. In this way, at the end of Algorithm 5.2 *dmu.result* will contain the decision maker's perspective.

Algorithm 5.2 DMU-Processing

Require: dmu.inputs $\neq NULL$
Require: dmu.DMmodel $\neq NULL$
Ensure: dmu.result

```
1: dmu.PreProcess(dmu.DMmodel,dmu.inputs)
2: dmu.DoFusion(dmu.DMmodel,dmu.inputs)        //Scoring Preferences
```

Bearing in mind that the processing in a DMU depends on the *fusionist* attribute to produce the *dmu.result* (DMU output), the *DMmodel* should be set up properly. For example, when LSP is used to produce the DMU output, the *DMmodel* should reflect the attribute tree, the characterization of preferences and the aggregation structure (Section 4.4.1).

By means of a decision-making fusion tree structure with DMU nodes, one could represent different multilevel organizational structures. In this way, the DMU at the highest hierarchical level will contain the fused information given by other DMU nodes located at the other levels in a hierarchy. Thus, the propagation of preferences is summarized as several DMU nodes requiring the client's preferences from its children until the lowest level is reached. At the lowest level, each DMU will process its inputs —i.e., the client's preferences and the decision maker's preferences— producing the desired output —i.e., a fusion of preferences that includes a score reflecting the decision maker's perspective.

Next section presents a detailed example of how the proposed model performs the fusion of preferences, given by potential customers and corporate managers, throughout a multilevel organizational structure of a company.

5.4 Illustrative Example

To illustrate how the preferences from different perspectives are fused using the proposed decision-making model, the following example is presented:

A decision about a brand-new model of 'winter shoe'. A shoe company has to decide about the proper combination of features (criteria) for a brand-new model of "winter shoe" (product). This company has a two-level organizational structure (Figure 5.9), where the *Intermediate Managers* are located at Level 1, and the *General Manager* of the company is located at Level 2.

To make the decision, the *General Manager* will take into account the point of view of two intermediate managers, namely the *Product Manager* and the

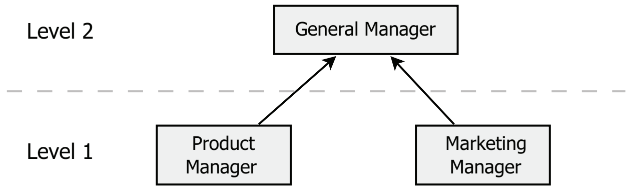


Figure 5.9: Organizational structure of a shoe company.

Marketing Manager. On the one hand the perspective of the *Product Manager* is mainly focused on the function of the new winter shoe (product), and to keep the business objectives set for this product. On the other hand, the perspective of the *Marketing Manager* is focused on the style of the brand-new shoe.

The *General Manager* has asked both managers to decide with respect to their competence area, specifically:

1. The *Product Manager* has to decide about the *water-resistant* feature of the shoe model (function of the product).
2. The *Marketing Manager* has to decide about the *style* of the shoe for a large number of potential customers.

To make their corresponding decisions, the managers require gathering some data that support what they consider to be important for the product.

The *Product Manager* considers that the *water-resistant* feature is based on: (1) the (main) *material* from which the upper part of the shoe is made, and (2) the (overall) *weight* including all the insulation, padding, and waterproofing that winter shoes need. Bearing in mind that the product is expected to satisfy a large number of customers (business objective), this manager decides to collect opinions (i.e., preferences) from potential customers for these criteria through social media. This manager considers that there are some opinions that are worthy of notice (i.e. noticeable opinions) such as the opinions from frequent clients. Using this kind of opinions, this manager evaluates the representativeness of grouped similar opinions regarding the weight (or material) for a winter-shoe. The computed representativeness values are used next to find out the preferred combination of material and weight. This combination results in his/her score regarding the 'water-resistant' feature of a winter-shoe. From his/her perspective, in this combination the representativeness regarding the 'weight' can be replaced by the representativeness regarding the 'material' and vice versa. However, their respective level of importance are 0.7 for the 'weight' and 0.3 for the 'material' criteria.

The *Marketing Manager* considers that the *style* is given by: (1) the *comfort*, and (2) the *modishness* (i.e., stylish in a modern way) of the brand-new shoe. This manager also decides to gather opinions (or preferences) given by potential customers for these criteria through different information sources.

This manager uses the total number of grouped similar opinions for evaluating the comfort representativeness for a winter-shoe. However, for evaluating the modishness representativeness, he/she also uses the number of noticeable opinions. The computed representativeness values are used next to find out the preferred combination of comfort and modishness. This combination results in his/her score regarding the 'style' feature of a winter-shoe. From his/her perspective, in this combination the representativeness regarding both the 'comfort' and the 'modishness' should be simultaneously satisfied and both aspects are equally important, i.e. the level of importance is 0.5 for each of them.

The *General Manager* will make his/her decision based on the relevant data provided by his/her subordinates, i.e., the intermediate managers. The data reported by the Product Manager includes the scores denoting the 'water-resistant' feature while the data reported by the Marketing Manager includes the scores regarding the style of the product. The General Manager considers that scores greater than 0.8 are acceptable, but he/she prefers scores greater than or equal to 0.9. These scores are used next to compute an overall score for the combination of features of a winter-shoe. From his/her perspective, in this combination the scores regarding both the 'water-resistant' and the 'style' should be simultaneously satisfied and are equally important, i.e. the level of importance is 0.5 for each of them.

5.4.1 Data Set Description

According to the example, both intermediate managers require to gather preferences from potential customers of the winter-shoe. Thus, these preferences have been simulated as follows:

Preferences collected by the Product Manager. For the *material* criterion, 1198 trapezoidal membership functions were randomly generated using a uniform distribution. Among these membership functions, 100 are considered to represent noticeable opinions. Each membership function was generated in the $[0\%, 100\%]$ domain representing the percentage of leather that a potential customer prefers in a winter-shoe. Likewise, for the *weight* criterion, 1498 trapezoidal membership functions were randomly generated (using a uniform distribution) limited to the [500,2500]-grams domain. In this case, 125 membership functions are considered to represent noticeable opinions. As an example, Figure 5.10 depicts a membership function μ_{weight} representing the opinion about the preferred *weight* of a winter-shoe given by a potential customer. It could be noticed that this customer prefers shoes having a weight around 2100 grams.

Preferences collected by the Marketing Manager. Regarding the *comfort* level, 798 membership functions were randomly generated (using a uniform distribution) in the [0%,100%] domain, where 67 of them represent noticeable opinions. These membership functions denote percentages, where 0% and 100% corresponds to the lowest and the highest

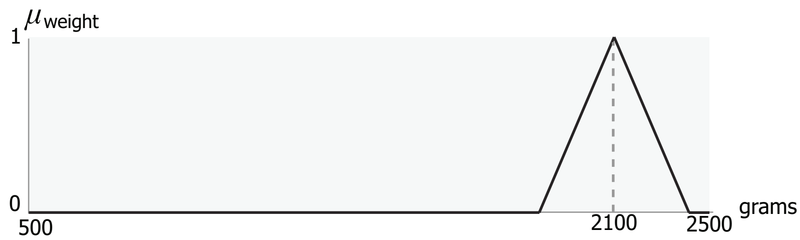


Figure 5.10: Preference regarding the *weight* of a winter-shoe given by a potential customer. Here, it is noticed that this customer prefers shoes having a weight around 2100 grams.

levels of comfort of a winter-shoe respectively. For the *modishness* criterion, 1298 membership functions denoting preferences were randomly generated as well. Among them, there are 109 membership functions representing noticeable opinions. These membership functions denote percentages, where 0% corresponds to a fully non-stylish shoe, and 100% corresponds to a fully stylish shoe.

In contrast to the previous data sets, the General Manager uses the data processed by the intermediate managers. So, data generation is not needed.

5.4.2 Modeling and Processing

To represent the organization given in the example, a structure with three DMU nodes is used. Since each manager acts as a decision maker at his/her corresponding DMU, he/she must provide his/her preferences (or constraints) by means of a decision model (DMmodel). Herein, all the DMmodels use LSP as a fusion method.

Recalling from Section 4.4.1 the steps of LSP, each *DMmodel* must include: (a) the attribute tree for aggregating preferences, (b) the characterization of preferences on the primary attributes, and (c) the aggregation structure. For the example, these components have been fixed as it will be detailed in the descriptions of the different DMUs.

In what follows, the decision model (DMmodel) and the processing (i.e., preprocessing of inputs and the fusion of preferences) in each DMU are presented.

5.4.2.1 Product Manager's DMU

DMmodel: The DMmodel of the Product Manager includes the following components.

(a) The *attribute tree* depicted in Figure 5.11 which is formed by two (primary) attributes, namely *material representativeness* and *weight representativeness*. Here, the term *representativeness* refers to the percentage of noticeable opinions in a cluster. For instance, the *material representativeness* refers to the percentage of noticeable opinions regarding the material of a winter-shoe in a cluster. The combination of the aforementioned two (primary) attributes results in the *water-resistant score*.

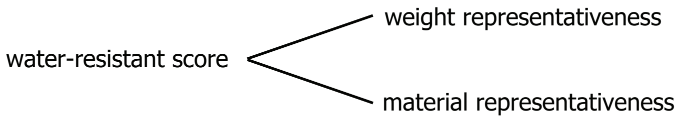


Figure 5.11: Attribute tree given by the Product Manager as a component of the Decision Model (DMmodel).

(b) The *characterization of preferences* regarding the representativeness over these attributes are given as follows. According to this manager, the representativeness level of a cluster of opinions is considered to be acceptable when at least 5% of the contained opinions are noticeable. However, this manager prefers cluster having 8% or more of noticeable opinions.

These preferences are given through membership functions $\mu_{\text{weight-rep}}$ and $\mu_{\text{material-rep}}$ representing the 'weight' and 'material' representativeness respectively. Figures 5.12a and 5.12b depict these preferences.

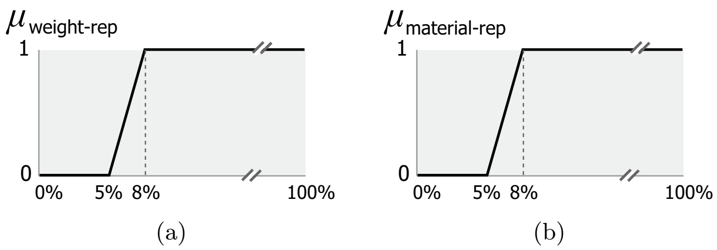


Figure 5.12: Characterization of preferences given by the Product Manager as a component of the Decision Model (DMmodel): (a) preferences for *weight representativeness* criterion, and (b) preferences for *material representativeness* criterion.

For instance, one may consider cluster $G_{\text{weight},0}$ with 264 opinions where 23 are noticeable (Figure 5.13). In this case, since the number of noticeable is greater than 21 (i.e., 8% of 264) the level of weight-representativeness is 1. It could be noticed in Figure 5.13 that, while the preference given by the Product Manager is about the 'weight representativeness', the preference given by a particular customer is about the 'weight' of a winter-shoe.

(c) The *aggregation structure* depicted in Figure 5.14 has been fixed based on the perspective of the Product Manager given in the ex-

Figure 5.13: A cluster of similar opinions regarding the *weight* of a winter-shoe, where the opinion of the particular customer of Figure fig:client-weight appears highlighted.

ample. This structure helps to compute the overall *water-resistant score*, $e_{\text{wr-score,j:k}}$, by aggregating the *weight representativeness* in cluster j, $e_{\text{weight-rep,j}}$, and the *material representativeness* in cluster k, $e_{\text{material-rep,k}}$ —here, while j refers to the identifier of a 'weight-cluster', k refers to the identifier of a 'material-cluster'. With this purpose, the structure uses the aggregation operator *D+-* denoting the level of replace-ability (or level of disjunction) between them given by $\omega = 0.8125$ and exponent $r = 5.802$. Moreover, the levels of relative importance associated to these criteria are $w_{\text{weight-rep}} = 0.7$ and $w_{\text{material-rep}} = 0.3$ respectively.

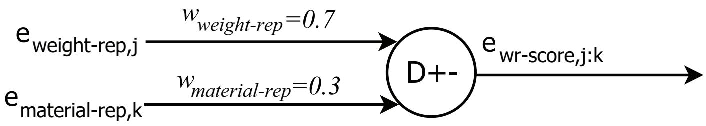


Figure 5.14: Aggregation structure given by the Product Manager as a component of the Decision Model (DMmodel). The GCD aggregator is denoted by the symbol *D+-* corresponding to the level of replace-ability given by $\omega = 0.8125$ and exponent $r = 5.802$ (Table 2.1).

Pre-processing of inputs: The *pre-processing of inputs* within this DMU reduces the number of collected data (1498 opinions regarding the shoe *weight* and 1198 opinions about the shoe *material*), through clusters of similar preferences (or opinions) for each criterion.

For the *weight* criterion, the data is reduced to 107 clusters. Here, each cluster j has two characteristics or attributes, namely the number of opinions (or preferences) denoted by $a_{p,j}$, and the number of noticeable opinions denoted by $a_{q,j}$. Figure 5.15 shows the *weight* criterion depicted as a rectangle C_{weight} which contains clusters of similar preferences (regarding this criterion) denoted by $G_{\text{weight,j}}$ where j corresponds to the cluster identifier.

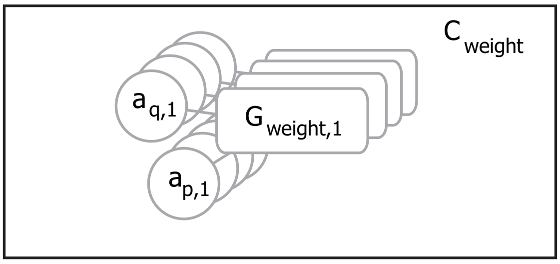


Figure 5.15: Potential customer's preferences over *weight* criterion expressed by clusters of similar preferences. Here, the *weight* criterion is depicted as a rectangle labeled C_{weight}, and each cluster j denoted by $G_{\text{weight,j}}$ has 2 attributes, namely the number of opinions denoted by $a_{p,j}$ and the number of noticeable opinions denoted by $a_{q,j}$.

For the *material* criterion, the data is reduced to 100 clusters. In the same way to the weight criterion, each cluster k has the same attributes that have been previously described. Analogously, to the previous criterion, each cluster k is denoted by $G_{\text{material,k}}$ depicted within a rectangle labeled C_{material} representing the criterion as shown in Figure 5.16.

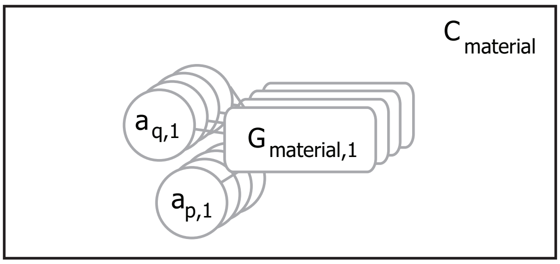


Figure 5.16: Potential customer's preferences over *material* criterion expressed by clusters of similar preferences. Here, the *material* criterion is depicted as a rectangle labeled C_{material}, and each cluster k denoted by $G_{\text{material,k}}$ has 2 attributes, namely $a_{p,j}$ and $a_{q,j}$.

It could be noticed that within this step the data is reduced down to about 10% using a shape-similarity approach. That is, from 1498 individual preferences to 107 clusters of similar opinions regarding the shoe *weight*, and from 1198 individual preferences to 100 clusters of similar opinions about the shoe *material*.

Fusion of preferences: The *fusion of preferences* within this DMU corresponds to the fusion of the weight-preferences with the material-preferences. This fusion results in an entity having the form

$$\{G_{\text{water-resistant score},[id]}\}_P^{\text{wr-score}},$$

where *id* is an identifier (of this fusion), wr-score $\in [0, 1]$ is an indicator reflecting to what extent the Product Manager prefers this fusion, and P identifies this manager.

For illustration purposes, one may consider two clusters: $G_{\text{weight},0}$ denoting preferences regarding the shoe weight (Figure 5.17), and $G_{\text{material},9}$ representing opinions about the shoe material (Figure 5.18) —here, the cluster identifiers are 0 and 9 respectively.

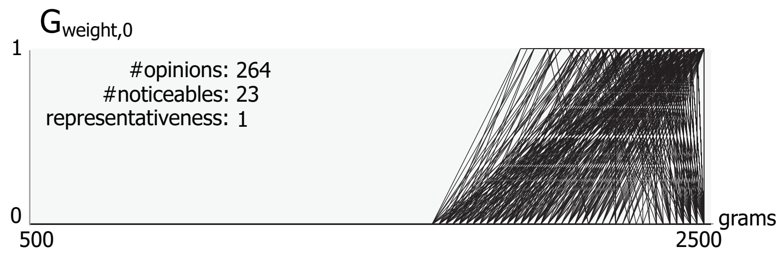


Figure 5.17: Cluster $G_{\text{weight},0}$ representing preferences about the shoe weight with attributes: number of opinions denoted by $a_{p,0}$, and number of noticeable opinions denoted by $a_{q,0}$.

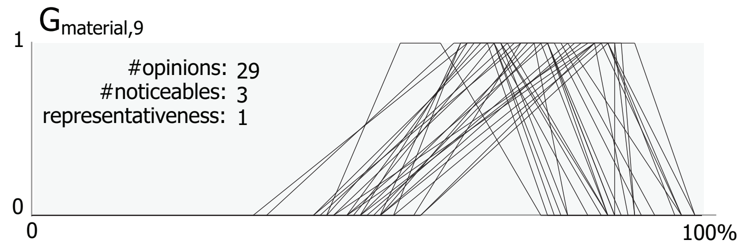


Figure 5.18: Cluster $G_{\text{material},9}$ representing opinions about the shoe material with attributes: number of opinions denoted by $a_{p,9}$ and number of noticeable opinions denoted by $a_{q,9}$.

Based on the Product Manager's *DMmodel*, the water-resistant score *wr-score* is computed as follows. First, the leaves of the attribute tree are evaluated, i.e., $e_{\text{weight-rep},0}$ and $e_{\text{material-rep},9}$ are computed:

$$\begin{aligned} e_{weight-rep,0} &= \mu_{weight-rep}\left(\frac{a_{q,0}}{a_{p,0}}\right) \\ &= \mu_{weight-rep}\left(\frac{23}{264}\right) \\ &= 1, \end{aligned}$$

and

$$\begin{aligned} e_{\text{material-rep},9} &= \mu_{\text{material-rep}}\left(\frac{a_{q,9}}{a_{p,0}}\right) \\ &= \mu_{\text{material-rep}}\left(\frac{3}{29}\right) \\ &= 1. \end{aligned}$$

Since these evaluations depend on the percentage of noticeable opinions, the corresponding percentages have been computed using $\frac{a_{q,j}}{a_{p,j}}$ for each cluster j.

Second, the *water-resistant score* is obtained using Equation 2.2 based on the aggregation structure (Figure 5.14). In this case the aggregation operator *D+-* is used, where $r = 5.802$ according to Table 2.1, and the levels of relative importance are $w_{\text{weight-rep}} = 0.7$ and $w_{\text{material-rep}} = 0.3$.

$$\begin{aligned} e_{\text{wr-score},0:9} &= (w_{weight-rep} \cdot e^r_{weight-rep,0} + w_{material-rep} \cdot e^r_{material-rep,9})^{\frac{1}{r}} \\ &= (0.7 \cdot 1^{5.802} + 0.3 \cdot 1^{5.802})^{\frac{1}{5.802}} \\ &= 1. \end{aligned}$$

The previously obtained value corresponds to the *water-resistant score* given by the Product Manager when clusters $G_{weight,0}$ and $G_{material,9}$ are used. Therefore, this fusion results in an entity denoted by

$$\{G_{water\text{-}resistant\ score,[w0:m9]}\}^1_P.$$

For tracking purposes, the identifier *w0:m9* has been assigned to this entity to indicate that it results from the fusion between cluster 0 for *weight* and cluster 9 for *material.*

In a similar way, the *water-resistant score* for each cluster combination regarding the *weight* and *material* representativeness can be performed. Recalling that in this example there are 107 clusters of preferences regarding the *weight* criterion and 100 clusters of preferences regarding the *material* criterion, the entity identifiers *wj:mk* for $j = 1, \ldots, 107$ and $k = 1, \ldots, 100$ range from *w1:m1* to *w107:m100.*

5.4.2.2 Marketing Manager's DMU

DMmodel: The DMmodel of the Marketing Manager includes the following components.

(a) The *attribute tree* depicted in Figure 5.19 which is formed by two (primary) attributes, namely *comfort representativeness* and *modishness representativeness.* The combination of these attributes results in the *style score.*

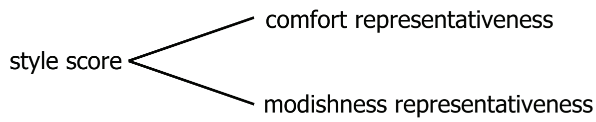


Figure 5.19: Attribute tree given by the Marketing Manager as a component of the Decision Model (DMmodel).

(b) The *characterization of preferences* regarding the representativeness over these attributes are given as follows. According to this manager, the *comfort representativeness* is considered to be acceptable when the number of opinions is at least 80. However, this manager prefers having 100 opinions or more in a cluster. Analogously, this manager's preferences regarding the *modishness representativeness* indicates that it is acceptable when the number of noticeable opinions are at least 5. However, this manager prefers 20 or more noticeable opinions.

These preferences are given by means of the membership functions $\mu_{\text{comfort-rep}}$ and $\mu_{\text{modishness-rep}}$ representing the manager's demands for 'comfort' and 'modishness' representativeness respectively. Figures 5.20a and 5.20b depict these preferences.

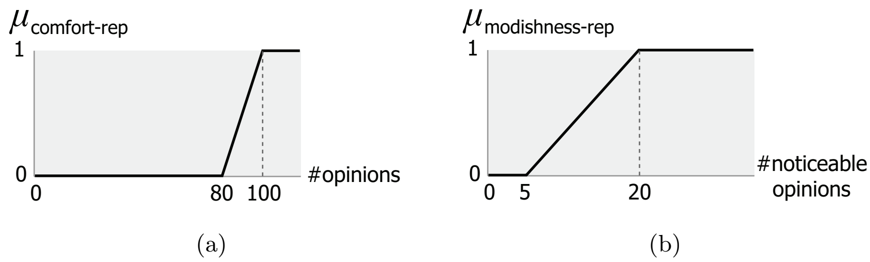


Figure 5.20: Characterization of preferences given by the Marketing Manager as a component of the Decision Model (DMmodel): (a) preferences for *comfort* criterion, and (b) preferences for *modishness* criterion.

(c) The *aggregation structure* depicted in Figure 5.21 has been fixed based on the perspective of the Marketing Manager given in the example. In this case, the *style score*, $e_{\text{style-score,j:k}}$ is computed by aggregating the *comfort representativeness* in cluster j, $e_{\text{comfort-rep},j}$, and the *modishness representativeness* in cluster k, $e_{\text{modishness-rep},k}$ —here, while j refers to the identifier of a 'comfort-cluster', k refers to the identifier of a 'modishness-cluster'. With this purpose, the structure uses the aggregation operator *C-+* denoting the level of of simultaneity (or level of conjunction) given by $\alpha = 0.6875$ and exponent $r = -0.148$ according to Table 2.1. Additionally, the levels of relative importance of the criteria have equal (weight) values, i.e.,

$$w_{\text{comfort-rep}} = w_{\text{modishness-rep}} = 0.5$$

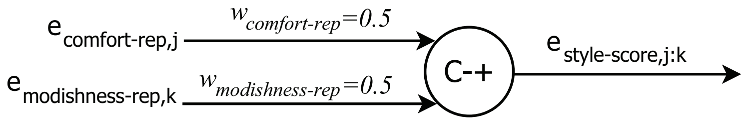


Figure 5.21: Aggregation structure given by the Marketing Manager as a component of the Decision Model (DMmodel). The GCD aggregator is denoted by the symbol *C-+* corresponding to the level of simultaneity given by $\alpha = 0.6875$ and exponent $r = -0.148$ (Table 2.1).

Pre-processing of inputs: In a similar way to the pre-processing of inputs of the Product Manager, within this DMU the number of collected data is reduced.

For the *comfort* criterion, the data is reduced from 798 individual opinions to 87 clusters. Here, each cluster j has two characteristics (or attributes), namely the number of opinions (or preferences) denoted by $a_{p,j}$, and the number of noticeable opinions denoted by $a_{q,j}$. Figure 5.22 shows the $comfort$ criterion depicted as a rectangle C_{comfort} which contains clusters of similar preferences (regarding this criterion) denoted by $G_{\text{comfort,j}}$ where j corresponds to the cluster identifier.

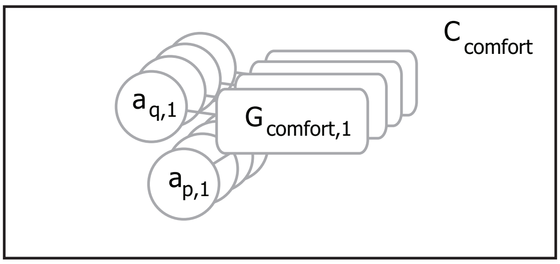


Figure 5.22: Potential customer's preferences over *comfort* criterion expressed by clusters of similar preferences. Here, the *comfort* criterion is depicted as a rectangle labeled C_{comfort}, and each cluster j denoted by $G_{\text{comfort,j}}$ has 2 attributes, namely $a_{p,j}$ and $a_{q,j}$.

For the *modishness* criterion, the data is reduced from 1298 individual preferences to 101 clusters. In the same way as in the comfort criterion, each cluster k has the same two attributes that have been previously described. Thus, each cluster k is denoted by $G_{\text{modishness,k}}$ depicted within a rectangle labeled $C_{\text{modishness}}$ representing the criterion as shown in Figure 5.23.

Fusion of preferences: The *fusion of preferences* within this DMU corresponds to the fusion of the comfort-preferences with the modishness-preferences. This fusion results in an entity having the form

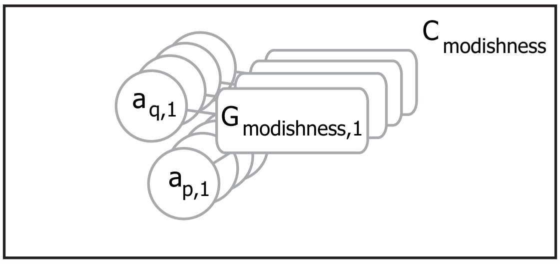


Figure 5.23: Potential customer's preferences over *modishness* criterion expressed by clusters of similar preferences. Here, the *modishness* criterion is depicted as a rectangle labeled $C_{\text{modishness}}$, and each cluster k denoted by $G_{\text{modishness,k}}$ has 2 attributes, namely $a_{p,k}$ and $a_{q,k}$.

$$\{G_{\text{style score},[id]}\}_{M}^{\text{style-score}},$$

where id is an identifier (of this fusion), style-score $\in [0,1]$ is an indicator reflecting to what extent the Marketing Manager prefers this fusion, and M identifies this manager.

For illustration purposes, one may consider two clusters: $G_{\text{comfort},0}$ denoting preferences regarding the shoe comfort (Figure 5.24), and cluster $G_{\text{modishness},6}$ representing opinions about the shoe modishness (Figure 5.25). So, here the cluster identifiers are 0 and 6 respectively.

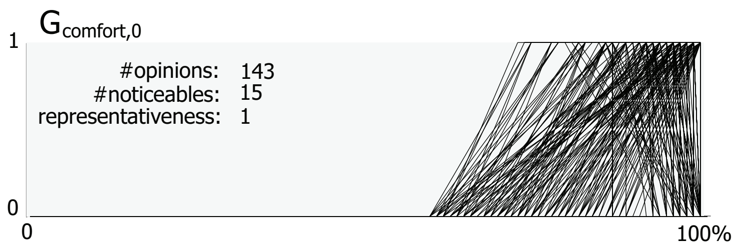


Figure 5.24: Cluster $G_{\text{comfort},0}$ representing preferences about the shoe comfort with attributes: number of preferences denoted by p, and number of frequent customers denoted by q.

Based on the Marketing Manager's *DMmodel*, the *style-score* is computed as follows. First, the leaves of the attribute tree are evaluated, i.e., $e_{\text{comfort-rep},0}$ and $e_{\text{modishness-rep},6}$ are computed:

$$\begin{aligned} e_{\text{comfort-rep},0} &= \mu_{\text{comfort-rep}}(a_{p,6}) \\ &= \mu_{\text{comfort-rep}}(123) \\ &= 1, \end{aligned}$$

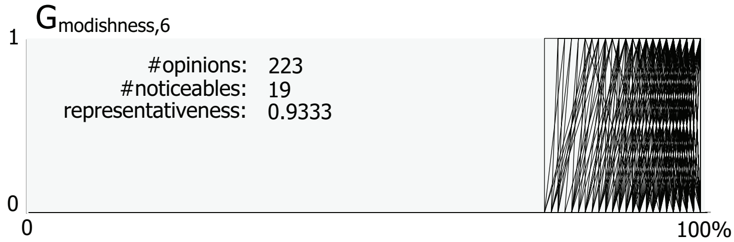


Figure 5.25: Cluster $G_{\text{modishness},6}$ representing preferences about the shoe modishness with attributes: number of preferences denoted by p, and number of frequent customers denoted by q.

and

$$\begin{aligned} e_{\text{modishness-rep},6} &= \mu_{\text{modishness-rep}}(a_{q,6}) \\ &= \mu_{\text{modishness-rep}}(19) \\ &= \frac{19-5}{20-5} \\ &= 0.93333. \end{aligned}$$

Since these evaluations depend on the percentage of noticeable opinions, the corresponding percentages have been computed using $\frac{a_{q,j}}{a_{p,j}}$ for each cluster j. It could be noticed that to evaluate $e_{\text{modishness-rep},6}$ from the the point of view of the Marketing Manager, this value is interpolated using $\mu_{\text{modishness-rep}}(x) = \frac{x-5}{20-5}$.

Second, the *style score* is obtained using Equation 2.2 based on the aggregation structure shown in Figure 5.21. Here, the aggregation operator C-+ is used and the weights $w_{\text{comfort-rep}} = w_{\text{modishness-rep}} = 0.5$ are given.

$$\begin{aligned} e_{\text{style-score}} &= (w_{\text{comfort-rep}} \cdot e^r_{\text{comfort-rep},0} \\ &\quad + w_{\text{modishness-rep}} \cdot e^r_{\text{modishness-rep},6})^{\frac{1}{r}} \\ &= (0.5 \cdot 1^{-0.148} + 0.5 \cdot 0.9333^{-0.148})^{\frac{1}{-0.148}} \\ &= 0.9660. \end{aligned}$$

The previously obtained value corresponds to the *style score* given by the Marketing Manager when clusters $G_{\text{comfort},0}$ and $G_{\text{modishness},6}$ are used. Therefore, this fusion results in an entity denoted by

$$\{G_{style\ score,[c0:m6]}\}^{0.9660}_{M}.$$

This value corresponds to the *score* given by the Marketing Manager for the shoe *style* when using clusters $G_{comfort,0}$ and $G_{modishness,6}$. For tracking purposes, the identifier $c0 : h6$ has been assigned to this entity to indicate that it results from the fusion of cluster 0 for *comfort* and cluster 6 for *modishness*.

5.4.2.3 General Manager's DMU

DMmodel: The DMmodel of the Product Manager includes the following components.

(a) The *attribute tree* depicted in Figure 5.26 which is formed by two (primary) attributes, namely *water-resistant score* and *style score*. The combination of these attributes results in the *winter-shoe score*.

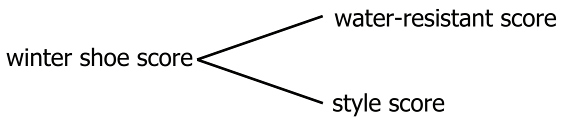


Figure 5.26: Attribute tree given by the General Manager as a component of the Decision Model (DMmodel).

(b) The *characterization of preferences* regarding the scores over these attributes are given as follows. According to the General Manager, the scores given by the intermediate managers are considered to be acceptable when these are at least 0.8. However, this manager prefers scores higher than or equal to 0.9.

These preferences are given through membership functions $\mu_{wr-score}$ and $\mu_{style-score}$representing the 'water-resistant' and 'style' scores respectively. Figures 5.27a and 5.27b depict these preferences.

Figure 5.27: Characterization of preferences given by the Marketing Manager as a component of the Decision Model (DMmodel): (a) preferences for *water-resistant score* criterion, and (b) preferences for *style score* criterion.

(c) The *aggregation structure* depicted in Figure 5.28 has been fixed based on the perspective of the General Manager given in the example. This structure helps to compute the *winter-shoe score*, given by

$e_{\text{shoe score,j:k}}$, through aggregating the *water-resistant score* in cluster j, $e_{\text{wr-score,j}}$, and the *style score* in cluster k, $e_{\text{style-score,k}}$. Here, j = 'w0:m9' refers to the entity identifier of

$$\{G_{\text{water-resistant score},[w0:m9]}\}_P^1,$$

and, k = 'c0:h6' refers to the entity identifier of

$$\{G_{\text{style score},[\text{c}0:h6]}\}_M^{0.9660}.$$

With this purpose, the structure uses the aggregation operator *C-+* denoting the level of simultaneity (or level of conjunction) given by $\alpha = 0.6875$ and $r = -0.148$ between criteria *water-resistant score* and *style score*. In this case, the levels of relative importance associated to the criteria are $w_{\text{wr-score}} = w_{\text{style-score}} = 0.5$.

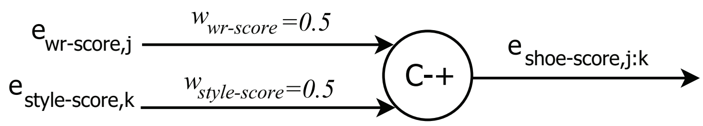


Figure 5.28: Aggregation structure given by the General Manager as a component of the Decision Model (DMmodel). The GCD aggregator is denoted by the symbol *C-+* corresponding to the level of simultaneity given by $\alpha = 0.6875$ and exponent $r = -0.148$ (Table 2.1).

Pre-processing of inputs: The *pre-processing of inputs* within this DMU reduces the data provided by both the Product Manager and the Marketing Manager. These data are reduced based on the strategy specified by the General Manager. For illustration purposes, one may consider that the strategy specified by the General Manager corresponds to the selection of the **"top scored entity"** given by each manager. Thus, the inputs of the General Manager's DMU are two entities, namely the top scored entity given by the Product Manager and the top scored entity given by the Marketing Manager. For illustration purposes, one may consider that these two entities are

$$\{G_{\text{water-resistant score},[w0:m9]}\}_P^1$$

and

$$\{G_{\text{style score},[\text{c0:h6}]}\}_M^{0.9660}$$

given by the Product Manager and the Marketing Manager respectively.

Fusion of preferences: The *fusion of preferences* within this DMU corresponds to the fusion of the aforementioned entities and results in an entity having the form

$$\{G_{\text{shoe score},[id]}\}_G^{\text{shoe-score}},$$

where *id* is an identifier (of this fusion), *shoe-score* $\in [0,1]$ is an indicator reflecting to what extent the General Manager prefers this fusion, and G identifies this manager.

Based on the General Manager's *DMmodel*, the *shoe-score* is computed. First, the leaves of the attribute tree are evaluated, i.e., $e_{wr\text{-}score,0}$ and $e_{style\text{-}score,6}$ are obtained:

$$\begin{aligned} e_{\text{wr-score,w0:m9}} &= \mu_{\text{wr-score}}(1) \\ &= 1 \end{aligned}$$

and

$$\begin{aligned} e_{\text{wr-score,c0:h6}} &= \mu_{\text{style-score}}(0.9660) \\ &= 1. \end{aligned}$$

Second, the *shoe score* is obtained using Equation 2.2 based on the aggregation structure (Figure 5.28). In this case the aggregation operator *C+-* is used, $\alpha = 0.6875$ and $r = -0.148$ according to Table 2.1, and the levels of relative importance are $w_{\text{wr-score}} = w_{\text{style-score}} = 0.5$.

$$\begin{aligned} e_{\text{shoe-score},[w0:m9][c0:h6]} &= (w_{\text{wr-score}} \cdot e^r_{\text{wr-score},w0:m9} \\ &\qquad + w_{\text{style-score}} \cdot e^r_{\text{style-score},c0:h6})^{\frac{1}{r}} \\ &= (0.5 \cdot 1^{-0.148} + 0.5 \cdot 1^{-0.148})^{\frac{1}{-0.148}} \\ &= 1. \end{aligned}$$

The previously obtained value corresponds to the *shoe score* given by the General Manager when the top entities provided by the intermediate managers are used. Therefore, this fusion results in an entity denoted by

$$\{G_{\text{shoe score},[w0:m9]:[c0:h6]}\}_G^1.$$

Considering that the proposed model uses a decision-making fusion tree structure with DMU nodes based on the company's organizational structure, the fusion of preferences at the General Manager's DMU provides a set of scored clusters representing the combination of features (criteria) for a brand-new model of winter shoe. In this example, where the "top" strategy was used, the fusion result consists in an entity with a score given by the General Manager reflecting his/her preferences over the decisions given by his/her subordinates, i.e. the intermediate managers of the company. This result allows the General Manager to decide about the combination of features of a winter-shoe based on the evaluation according to his/her DMmodel.

It could be noticed that the new entity identifier facilitates to track different entities that are involved during the scoring of the criteria (i.e., weight, material, comfort and modishness) through different organizational levels. In this way, it might be also possible to involve additional attributes in the scoring that are available at lower levels of the DMU-tree structure.

5.5 Conclusions

This chapter proposed an original decision-making model that performs a fusion of preferences from different perspectives over multiple criteria. This model is suitable for different (multilevel) organizational structures, where multiple decision makers contribute to some extent to make a final decision. Moreover, it handles preferences given by a large number of people with different backgrounds (i.e., different levels of knowledge, areas of expertise and personal profiles) where these preferences might be gathered through several sources such as fan pages, surveys, polls and social network applications.

In a similar way that an organization delegates decisions to different managers, the proposed decision-making model allows a decision maker to ask preferences to its subordinates with respect to their competence areas —For instance, a General Manager asking other (intermediate) managers their preferences from an economical, technical, and environmental perspective. To manage the preferences from each perspective, the novel concept of a *decision-making unit (DMU)* has been introduced as a component of the proposed decision-making model.

The processing in a DMU has two steps, namely the *pre-processing of inputs* and the *fusion of preferences*. During the *pre-processing of inputs*, the client's preferences are adapted for their further processing, while the *fusion of preferences* allows for the scoring of preferences according to a decision maker's perspective. Thus, the processing in a DMU consists in the fusion of the preferences (or constraints) given by a single decision maker with others received as inputs (e.g., given by other DMUs or potential customers). Moreover, a tree structure with DMU nodes allows for the propagation of preferences throughout an organizational structure. In this dissertation, the propagation of preferences consists in enriching information (at the level) where a final decision is made. In this way, the decision maker at the highest organizational level could make a more informed decision.

Further work using this model aims to compare the results when different multicriteria methods are applied during the fusion step. An interesting topic that requires future research is the uncertainty propagation and its impact in the results, specially when a large amount of information might be gathered through different sources including social media.

References

[1] Thomas L. Saaty. *The Analytic Hierarchy Process.* McGraw-Hill, New York, 1980.

[2] Ching-Lai Hwang and Kwangsun Yoon. *Multiple attribute decision making: methods and applications a state-of-the-art survey*, volume 186. Springer Science & Business Media, 2012.

[3] Thomas L. Saaty. *What is the analytic hierarchy process?* Springer, 1988.

[4] Chi-Cheng Huang, Pin-Yu Chu, and Yu-Hsiu Chiang. *A fuzzy {AHP} application in government-sponsored R&D project selection.* Omega, 36(6):1038 – 1052, 2008. A Special Issue Dedicated to the 2008 Beijing Olympic Games.

[5] Witold Pedrycz. *Granular Computing: Analysis and Design of Intelligent Systems*, volume 13. CRC Press/Francis Taylor, 2013.

[6] Ronald R. Yager and Antoine Kelman. *An extension of the analytical hierarchy process using OWA operators.* Journal of Intelligent & Fuzzy Systems, 7(4):401–417, 1999.

[7] Alessio Ishizaka and Ashraf Labib. *Review of the main developments in the analytic hierarchy process.* Expert Systems with Applications, 38(11):14336–14345, 2011.

[8] Petra Grošelj, Lidija Zadnik Stirn, Nadir Ayrilmis, and Manja Kitek Kuzman. *Comparison of some aggregation techniques using group analytic hierarchy process.* Expert Systems with Applications, 42(4):2198–2204, 2015.

[9] Jie Lu, Jun Ma, Guangquan Zhang, Yijun Zhu, Xianyi Zeng, and Ludovic Koehl. *Theme-based comprehensive evaluation in new product development using fuzzy hierarchical criteria group decision-making method.* Industrial Electronics, IEEE Transactions on, 58(6):2236–2246, 2011.

[10] Robert H. Ashton. *Combining the judgments of experts: How many and which ones?* Organizational Behavior and Human Decision Processes, 38(3):405–414, 1986.

[11] David Blackwell and Lester Dubins. *Merging of opinions with increasing information.* The Annals of Mathematical Statistics, pages 882–886, 1962.

[12] James K. Hammitt and Yifan Zhang. *Combining experts judgments: comparison of algorithmic methods using synthetic data.* Risk Analysis, 33(1):109–120, 2013.

[13] Miguel Delgado, Francisco Herrera, Enrique Herrera-Viedma, and Luis Martínez. *Combining numerical and linguistic information in group decision making.* Information Sciences, 107:177–194, 1998.

[14] Witold Pedrycz, Petr Ekel, and Roberta Parreiras. *Fuzzy Multicriteria Decision-Making: Models, Methods and Applications.* John Wiley and Sons, Chichester, UK., 2011.

[15] Ana Tapia-Rosero, Antoon Bronselaer, and Guy De Tré. *A method based on shape-similarity for detecting similar opinions in group decision-making.* Information Sciences, 258:291–311, 2014.

[16] Jozo Dujmović and Guy De Tré. *Multicriteria methods and logic aggregation in suitability maps.* International Journal of Intelligent Systems, 26(10):971–1001, 2011.

Chapter 6

Main Contributions and Further Research

This chapter summarizes the main contributions of the conducted research. Additionally, this chapter presents interesting topics that need further research based on these contributions.

6.1 Main Contributions

This dissertation presents a novel decision-making model that performs the fusion of preferences from different perspectives within a multilevel organizational structure. These preferences are considered to be provided by a large number of participants, as well as provided by multiple decision makers. To that end, this dissertation provides several contributions which are detailed as follows.

6.1.1 Handling a Large Number of Preferences

The main contribution while answering the question

(1) how to simplify the complexity of a decision-making problem that involves the preferences given by a community?

is a novel *shape-similarity detection method*. The aim of this method is to reduce the complexity of handling a large number of preferences by grouping them by similarity. To that end, each preference is represented by means of a membership function and, then, it is annotated using a shape-symbolic notation. The obtained shape-symbolic notations are used in a new shape-similarity measure to compute their similarity within a clustering process. To the best of our knowledge, this is the first detection method based on the shape-similarity of membership functions that has been presented in the literature.

Along with the shape-similarity detection method, the novel contributions are

(i) a *shape-symbolic notation*, which is a novel representation of a membership function used to depict its shape characteristics;

(ii) the algorithms *getShapeString* and *getFeatureString*, which are used to build a shape-symbolic notation from piecewise linear functions such as trapezoidal membership functions; and

(iii) a *shape-similarity measure*, which facilitates the comparison among membership functions that have been annotated by means of shape-symbolic notations.

Here, it is worth to mention that an inner novelty of the shape-similarity measure is a technique, based on the *edit (or Levenshtein) distance*, to compute the edit distance between two shape-symbolic notations.

6.1.2 Identifying and Evaluating Relevant Opinions

The main contribution to address the research question

(2) how to identify and evaluate preferences considered being relevant from a decision maker's perspective?

is a *methodology whereby groups of preferences (or opinions) are evaluated in order to determine their relevance according to a decision maker's perspective.* The novel contributions within this methodology are

(i) a *model for aggregating preferences on group attributes*;

(ii) a *cohesion measure* as a characteristic (or attribute) present in groups of preferences formed by means of the shape-similarity detection method;

(iii) *two computational methods*, one based on an *extended shape-symbolic notation* and the other based on a *geometric approach*, to obtain the cohesion measure of a group; and

(iv) a *procedure to compute the relevance of a group* based on the aforementioned aggregation model.

6.1.3 Fusion of Preferences from Different Perspectives

The main contribution to answer the research question

(3) How to simplify the complexity of a decision-making problem that includes multiple perspectives in an organizational environment, as well as the preferences given by a large number of persons?

is a novel *decision-making fusion tree model* that performs the fusion of preferences from different perspectives in a multilevel organizational structure. Novel contributions within this topic are

(i) a *decision-making unit (DMU)*, which is an abstract representation of a decision-making problem that allows for the fusion of preferences provided by a single decision-maker with the ones received as inputs. The preferences received as inputs in a DMU might come from a large group of persons involved in a decision, as well as other *DMUs* in a hierarchical structure;

(ii) a *schema for the propagation of preferences*, which allows a DMU to propagate the preferences throughout an organizational structure. In this way, the preferences are enriched while are propagated up to the level where a final decision is made.

The aforementioned contributions resulted in the publications of 2 published peer reviewed journal papers A1 ([1, 2]), 2 published book chapters B2 ([3, 4]), and 2 papers published in the proceedings of international conferences C1 ([5, 6]). Moreover, one of these publications ([6]) has received a Best Student Paper Award. In all of these publications, the author of this dissertation has contributed as first author.

6.2 Further Research

There are several interesting topics that require further research regarding the complexity of a decision-making problem. In particular, decision problems that are considered to be complex in the presence of a large number of fuzzy preferences and multiple perspectives are worthy of notice from the author's point of view. In this respect, a short list of topics that require further research is presented next.

- *Extensions of shape-symbolic notations for different types of fuzzy sets.* Considering that a shape-symbolic notation is a membership function's representation and within this dissertation an extended shape-symbolic notation allows for representing interval-valued fuzzy sets, it could be interesting to study other extensions for representing different types of fuzzy sets —e.g. Atanassov intuitionistic fuzzy sets, hesitant fuzzy sets, type-2 fuzzy sets, among others that can be used for representing preferences.

- *Granularity on Shape-Symbolic Notations.* An interesting subject to further study is the effect on the accuracy in a membership function's representation when using different levels of granularity on the linguistic terms that are used to build the shape-symbolic notation.

- *Identification of additional attributes on groups of similar preferences.* Bearing in mind the availability of different methods in the computational intelligence area, the study of additional attributes within a group might contribute to a better discrimination of groups regarding their relevance. For example, in consensual processes an attribute reflecting the group's attitude toward consensus can be considered [7].

- *Comparison of the proposed fusion model when using different aggregation operators.* Although the selection of an aggregation operator usually depends on its properties and the context where the operator is going to be used, it could be interesting to make a comparison of the usability of the fusion model presented in this dissertation when using different carefully-selected aggregation operators.

- *Propagation of uncertainty.* Considering that uncertainty can stem from different sources, handling the uncertainty within a DMU could help to reach an even more informed decision.

- *Applicability of the proposed model to real-world applications.* For instance, decision-making problems regarding natural resources, such as water management [8], where a community and different stakeholders are involved.

The aforementioned list is far from being comprehensive, but can be used as a starting point to build on different components proposed within this dissertation as well as the applicability of the proposed model to real-world applications.

References

[1] Ana Tapia-Rosero, Antoon Bronselaer, and Guy De Tré. *A method based on shape-similarity for detecting similar opinions in group decision-making.* Information Sciences, 258:291–311, 2014.

[2] Ana Tapia-Rosero, Antoon Bronselaer, Robin De Mol, and Guy De Tré. *Fusion of preferences from different perspectives in a decision-making context.* Information Fusion, 29:120–131, 2016.

[3] Ana Tapia-Rosero and Guy De Tré. *A Cohesion Measure for Expert Preferences in Group Decision-Making.* In Krassimir Atanassov, Władysław Homenda, Olgierd Hryniewicz, Janusz Kacprzyk, Maciej Krawczak, Zbigniew Nahorski, Eulalia Szmidt, and Sławomir Zadrozny, editors, New Trends in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics, volume II: Applications, pages 125–142. Systems Research Institute Polish Academy of Sciences, 2013.

[4] Ana Tapia-Rosero, Robin De Mol, and Guy De Tré. *Handling uncertainty degrees in the evaluation of relevant opinions within a large group.* In Juan Julian Merelo, Rosa Agostinho, José M. Cadenas, António Dourado, Kurosh Madani, and Filipe Joaquim, editors, Computational intelligence, volume 620 of *Studies in Computational Intelligence*, pages 283–299. Springer International Publishing, 2016.

[5] Ana Tapia-Rosero, Antoon Bronselaer, and Guy De Tré. *Similarity of membership functions: a shaped based approach.* In Proceedings of the 4th International Joint Conference on Computational Intelligence, pages 402–409. Ghent University, Department of Telecommunications and information processing, 2012.

[6] Ana Tapia-Rosero and Guy De Tré. *Evaluating relevant opinions within a large group.* In 6th International Conference on Fuzzy Computation Theory and Applications, pages 76–86. SciTePress, 2014.

[7] Iván Palomares, Rosa M. Rodríguez, and Luis Martínez. *An attitude-driven web consensus support system for heterogeneous group decision making.* Expert Systems with Applications, 40(1):139–149, jan 2013.

[8] Piero Fraternali, Andrea Castelletti, Rodolfo Soncini-Sessa, Carmen Vaca Ruiz, and Andrea Emilio Rizzoli. *Putting humans in the loop: Social computing for Water Resources Management.* Environmental Modelling & Software, 37:68–77, 2012.

Conclusions

This dissertation studies the complexity of decision making in the presence of a large number of heterogeneous fuzzy preferences and multiple perspectives. As a main result of this study, an original decision-making model is provided.

The new decision-making model can handle preferences that may be obtained from different sources (e.g., fan pages, surveys, polls or other social media applications where users can express their preferences) and it can handle multiple perspectives (e.g., social, technical, financial or environmental). As an advantage, this model considers that persons who express their preferences without deciding on the result (i.e., regular participants) and decision makers are considered to be participants in a decision-making process.

The model can be applied to help decision makers to reach better motivated decisions due to the diversity of the participants (i.e., participants with different levels of education, experience and areas of expertise). Moreover, the model performs the fusion of multiple perspectives while diminishes information loss based on a hierarchical structure. Thus, the model is suitable for different multilevel organizational structures enriching the information at the level where a final decision is made.

Besides a novel decision-making model, this dissertation has raised some interesting open issues. Although these are not addressed in this dissertation, these are presented next because these could further improve the applicability of the presented decision-making model.

One of them is the automatic characterization of preferences as fuzzy sets. This dissertation has presented how it is possible to model preferences by specifying either an ideal value, a range of values or by using a linguistic approach. However, these days it is desired to automatically represent preferences that are stated in reviews posted on social media.

Another interesting topic corresponds to the study of the calibration of preferences to facilitate the use of well-known concepts within a context. One may consider the term 'low temperature' as an example, where this term can be used in different climate zones. However, the meaning of this term for each member of a community may be different.

An open issue regarding the fact that the model can provide a solution set is related to the generation of different decision scenarios. Hence, optimization problems, strategic planning and suitability analysis can be considered topics for future research.

One topic that might improve the accuracy of the model in real-world applications is the inclusion of a feedback mechanism. This feedback mechanism might provide a bidirectional flow of preferences or constraints among the levels of a hierarchical structure.

It can be interesting to explore the application of the model in other research fields. For instance, in fields such as data mining and machine learning where recommender systems explore the prediction of preferences.

Many other issues remain open, however the aforementioned topics may be considered as a start to improve the purpose of this PhD study and to make more contributions to the field.