Quality in Crowdsourced Experience-Based Evaluations: Handling Subjective Responses

Marcelo Loor Romero

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

UNIVERSITEIT GENT

Vakgroep Telecommunicatie en Informatieverwerking
Voorzitter: prof. dr. ir. H. Bruneel
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2017 - 2018

ISBN 978-94-6355-106-9
NUR 982, 984
Wettelijk depot: D/2018/10.500/24

Examination Board

Supervisor:

Prof. Guy De Tré
Ghent University, Ghent, Belgium

Voting Members:

Prof. Filip De Turck (academic secretary), chair
Ghent University, Ghent, Belgium

Prof. Antoon Bronselaer
Ghent University, Ghent, Belgium

Prof. Sidharta Gautama, secretary
Ghent University, Ghent, Belgium

Prof. Peter Lambert
Ghent University, Ghent, Belgium

Prof. Marek Reformat
University of Alberta, Edmonton, Canada

Prof. Evdokia Sotirova
Prof. Dr. Assen Zlatarov University, Burgas, Bulgaria

GHENT UNIVERSITY

FACULTY OF ENGINEERING AND ARCHITECTURE

To "*Amorcito*" and "*Mijito*" ... my teammates

Acknowledgements

A heterogeneous group of important people deserves my deep gratitude for being directly or indirectly involved in the preparation of this dissertation.

To begin with, I would like to thank Guy, who I consider more than a mentor, a friend. He has advised me not only on scientific ideas, but also on good food, pre- and after-dinner drinks, nice places and other interesting topics. I really appreciate them all.

I am grateful for the support of the jury members, Prof. Bronselaer, Prof. Gautama, Prof. Lambert, Prof. Reformat and Prof. Sotirova, whose comments and suggestions made this dissertation more fair and readable. I am also grateful for the insights about intuitionistic fuzzy sets shared by Prof. Atanassov during a few conversations with him. His recommendations help me to move my thoughts on this kind of fuzzy sets forward.

Many thanks to my colleagues and friends at ESPOL, particularly to Katherine and Carlos for their convincing letters of recommendation. Also, many thanks to the DDCM team, Antoon, Christophe, Robin, Joachim, Hannah, Rashid and Toon, for sharing diverse and interesting experiences. Special thanks to Sylvia, Patrick, Davy and Philippe for your immediate assistance in technical, logistic or administrative aspects I had to cope with.

I owe my family and friends a debt of thanks. Gracias infinitas a mi esposa y a mi hijo, por ser parte del mejor equipo que uno podría desear. Gracias a mis padres por enseñarme a valorar lo importante y a creer que todo es posible. Gracias a mis hermanos por su ayuda en todo momento, y en especial a mi pequeña gran hermana por hacerse cargo de innumerables cosas durante mis estudios. Gracias a Ale y a Leo por mostrarme que es posible enfrentar duras pruebas y seguir adelante. Gracias a Amelita, Cami, Lu, Saskita, Julius, Fiore, Nico, Vero, George, David, Chío y Eddy, quienes suelen autodenominarse *"conejillos"* por degustar nuevas recetas. Gracias a la familia Tapia y en especial al Tío y a la Ñañi por su apoyo en todo momento. Many thanks to Claudine, Yana, Inge, Andy, Lucas, Matthias and Hanne for the nice talks and amazing moments and, in particular, for "adopting" Marce sometimes.

Last but not least, I acknowledge the support of the *careful reader*, whose comments and suggestions have always been and will be welcome.

Samenvatting

"Experience-based evaluations" of XBE's zijn beoordelingen die gebaseerd zijn op wat iemand door ervaring over een onderwerp verstaat of heeft geleerd. XBE's kunnen resulteren uit een expliciet evaluatieverzoek of kunnen bijvoorbeeld het gevolg zijn van ongevraagde opinies die op sociale media zijn geplaatst. Zo is de uitdrukking "*luchtvaartmaatschappij XYZ* kan ietwat worden beschouwd als *een betaalbare maatschappij* omdat er geen extra kosten gevraagd worden voor een tweede bagagestuk" bijvoorbeeld een XBE die gegeven werd door een persoon die antwoordde op het evaluatieverzoek "evalueer tot op welk niveau *luchtvaartmaatschappij XYZ* kan worden beschouwd als *een betaalbare maatschappij*", terwijl de sociale mediapost "*@LuchtvaartmaatschappijXYZ* is een *#BetaalbareLuchtvaartmaatschappij* wegens haar aanbiedingen :)" kan worden beschouwd als een XBE die het gevolg is van een ongevraagde opinie over *luchtvaartmaatschappij XYZ* die werd geplaatst door een persoon die een mooie aanbieding kreeg.

Hoewel XBE's sterk subjectief, onnauwkeurig, divers en mogelijks gekenmerkt door aarzeling kunnen zijn, kan de informatie die eruit geëxtraheerd wordt bedrijven en organisaties tastbare voordelen opleveren. Om die reden is het bij modern informatiebeheer belangrijk om op de juiste om te gaan met XBE's. Niettemin kan het behandelen van XBE's verschillende uitdagende taken omvatten. Dit is in het bijzonder het geval wanneer potentiële datakwaliteitsproblemen, zoals een gebrek aan betrouwbaarheid bij XBE's die afkomstig zijn van een grote en heterogene groep van (anonieme) bronnen, moeten worden behandeld.

Een adequate representatie is nodig om XBE's, die afkomstig zijn van personen die een ander begrip kunnen hebben van het concept dat wordt gevalueerd, te registreren en verwerken. Vandaar dat het vinden van een gepaste representatie een eerste uitdaging vormt voor het behandelen van XBE's. Voor deze taak worden in dit werk twee benaderingen voorgesteld: een benadering waarbij niets over de *context* van een XBE, d.i. de condities die in acht genomen werden wanneer de beoordeling werd gemaakt, wordt geregistreerd; en een andere benadering waarbij indicaties over de context worden geregistreerd. In dit verband worden bestaande en nieuwe modellen en concepten uit het gebied van de "computational intelligence", die erop gericht zijn om benaderende, haalbare en robuuste oplossingen te vinden, bestudeerd en voorgesteld in dit proefschrift. Meer specifiek wordt in dit proefschrift bestudeerd of en hoe concepten

zoals *vaagverzamelingen*, *intuïtionistische vaagverzamelingen*, *Pythagorische vaagverzamelingen* en *bipolaire voldoeningsgraden*, die alle verband houden met vaagverzamelingenleer, kunnen worden gebruikt om XBE's te modelleren. Onder deze concepten blijken *intuïtionistische vaagverzamelingen* voldoende en praktisch te zijn om subjectieve, onnauwkeurige en door aarzeling gekenmerkte XBE's te karakteriseren, die gegeven zijn door respondenten die eenzelfde *'mentaal beeld'* of begrip hebben van het concept dat wordt geëvalueerd. In dat geval moet de context van de XBE's niet worden geregistreerd omdat impliciet wordt verondersteld dat alle respondenten zich voor hun evaluatie hebben gebaseerd op dezelfde (of zeer soortgelijke) kenmerken van de geëvalueerde objecten. In andere gevallen waarbij XBE's gegeven worden door respondenten met een verschillende focus, is het niet mogelijk om aan te nemen dat dezelfde (of zeer soortgelijke) kenmerken zijn gebruikt. Het is dan handig om hints over hun beoordelingen te registreren. Voor dergelijke gevallen blijken "*augmented appraisal degrees*" (AAD's) en *"augmented (Atanassov) intuitionistisc fuzzy sets"* (AAIFS's), nieuwigheden in dit werk, meer geschikt om XBE's weer te geven.

Een tweede uitdaging heeft betrekking op de methoden en technieken die noodzakelijk zijn om XBE's te verwerken. Deze uitdaging gaat samen met de karakterisering van XBE's omdat het resultaat dat bekomen wordt na het verwerken van twee XBE's afhankelijk zal zijn van de manier waarop deze XBE's worden gerepresenteerd. Veronderstel bijvoorbeeld dat de uitdrukking "*luchtvaartmaatschappij XYZ* kan ietwat worden beschouwd als *een betaalbare maatschappij* omwille van haar terugkerende aanbiedingen" een XBE is die gegeven is door een andere persoon die voornoemd evaluatieverzoek beantwoordde. Als iemand enkel de niveaus registreert die binnen de XBE's in antwoord op het evaluatieverzoek werden aangegeven en deze vergelijkt, dan zullen beide XBE's overeenkomen omdat voor beide hetzelfde niveau '*ietwat betaalbare maatschappij'* geregistreerd werd. Anderzijds, indien ook de vermelde redenen worden geregistreerd, zal de vergelijking van deze XBE's een ander resultaat opleveren omdat de karakteriseringen 'ietwat betaalbare maatschappij'@'geen extra kost voor tweede bagagestuk' en 'ietwat betaalbare maatschappij'@'terugkerende aanbiedingen' verschillend zijn. Voor XBE's die gekarakteriseerd worden als elementen van een intuïtionistische vaagverzameling (IFS) worden bestaande en nieuwe similariteitsmaten voorgesteld en gebruikt om XBE's te vergelijken binnen het IFS-raamwerk. Verder wordt een nieuw softwarepakket, *IFSMetrics* genoemd, voorgesteld en beschreven waarmee zulke similariteitsmaten kunnen worden getest via het vergelijken van IFS's, die gesimuleerde XBE's karakteriseren die verkregen werden van een heterogene groep van respondenten. De resultaten van een empirische studie die werd uitgevoerd met dit softwarepakket suggereren dat similariteitsmaten die werden aangevuld met een soort van voetspoor van de vergelijking, betrouwbare vergelijkingen van dergelijke gesimuleerde XBE's opleveren. Voor XBE's die gekarakteriseerd worden als elementen van een AAIFS, is een nieuw raamwerk met nieuwe operatoren en methoden voorgesteld, waarvan wordt aangetoond dat het geschikt is om dergelijke XBE's betrouwbaar te vergelijken.

Een derde uitdaging is gerelateerd aan de *kwaliteit van XBE's*. XBE's van *hoge kwaliteit* worden in dit werk geacht XBE's te zijn, die geschikt zijn voor gebruik door een aanvrager. Dit betekent dat *bruikbaarheid* en *de mate waarin een XBE geschikt is voor gebruik*, worden beschouwd als belangrijke aspecten van de kwaliteit van XBE's. Omdat de *relevantie* van een XBE sterk gelinkt is aan zijn bruikbaarheid, heeft dit kenmerk speciale aandacht gekregen bij de kwaliteitsbeoordeling van XBE's. In dit verband wordt beschouwd dat de relevantie (en dus de kwaliteit) van de XBE's van een respondent afhangt van hoe goed er overeenstemming is tussen het begrip dat de respondent heeft over het geëvalueerde concept en het begrip dat de evaluatieaanvrager hierover heeft. Zowel nieuwe similariteitsmaten voor het IFS-raamwerk, als nieuwe vergelijkingsoperatoren en similariteitsmaten voor het nieuwe, uitgebreidere AAIFS-raamwerk worden voorgesteld en hun geschiktheid voor het bepalen van de mate waarin XBE's relevant zijn volgens een bepaald begrip, wordt aangetoond. Vandaar dat beide raamwerken worden beschouwd als haalbare opties om de volgens een bepaald begrip waargenomen kwaliteit van XBE's te meten.

Bijdragen tot het oplossen van deze drie geïdentificeerde uitdagingen vormt de doelstelling van dit proefschrift, die er concreter uit bestaat om te bestuderen of en hoe nieuwe concepten en methoden uit het gebied van de "computational intelligence" kunnen worden gebruikt om XBE's zodanig te karakteriseren en verwerken dat men adequaat kan omgaan met kwaliteitsaspecten van subjectieve gegevens die afkomstig zijn van een grote, heterogene groep respondenten. Een belangrijke bijdrage van dit werk, bekomen tijdens het zoeken naar oplossingen voor de geïdentificeerde uitdagingen, is de definitie van het nieuw concept "*Augmented Appraisal Degree*" (AAD). Zoals hierboven werd aangegeven zijn AADs voorgesteld voor de karakterisering van subjectieve, onnauwkeurige, diverse en mogelijks door aarzeling gekenmerkte XBE's, op een zodanige manier dat ze zich gemakkelijk lenen voor berekeningen. Dit concept vormt samen met andere nieuwe concepten en methoden het hiervoor aangehaalde uitgebreide AAIFS-raamwerk, dat het zowel toelaat om betrouwbare vergelijkingen te maken tussen XBE's die gegeven zijn door een heterogene groep van respondenten, als om de volgens een bepaald begrip waargenomen kwaliteit van XBE's te meten.

IFSMetrics is een andere belangrijke bijdrage. Zoals reeds werd vermeld, implementeert dit open-source softwarepakket een innovatieve experimentele studie die werd voorgesteld voor het testen van similariteitsmaten met IFS's die gesimuleerde XBE's karakteriseren die gegeven zijn door een heterogene groep van respondenten. A significant aspect van *IFSMetrics* is dat een onderzoeker of gebruiker de code kan gebruiken, aanpassen of uitbreiden om andere bestaande of nieuwe similariteitsmaten te testen.

Een bijkomende bijdrage van dit proefschrift is een nieuwe methode, de *"k-well-(un)fitted specimens"-methode* genoemd, die is voorgesteld om een benadering te berekenen van de mate waarin de contexten van XBE's over sociale media inhoud als gelijk worden waargenomen. Deze methode maakt enkel gebruik van de beoordelingsgraden die bevat zijn in de XBE's van een aantal specifieke sociale mediaberichten die door de aanvrager als relevant worden beschouwd voor het concept dat wordt bestudeerd. Door middel van een empirische studie met gesimuleerde XBE's werden voldoende aanwijzingen gevonden om te besluiten dat de "k-well-(un)fitted specimens"-methode geschikt is voor het behandelen van XBE's waarvoor enkel beoordelingsgraden beschikbaar zijn en voor het identificeren van mogelijke verschillen in het begrip dat respondenten uit een heterogene groep kunnen hebben over het concept dat wordt geëvalueerd. Een nieuwe methode, de *"post digest"-methode* genoemd, is een andere bijdrage in dit opzicht. Berichten op sociale media kunnen met deze methode worden verwerkt tot AAIFS's die XBE's karakteriseren. Men kan dan de faciliteiten van het uitgebreide AAIFS-raamwerk gebruiken om opinies (of berichten) te detecteren die gepost zijn door mensen die een gelijk begrip delen over een gegeven feit of onderwerp. De toepasbaarheid van de "post digest"-methode is geïllustreerd door middel van een voorbeeld waarbij beoordelingen van muziekalbums worden verwerkt om beoordelaars te vinden die een gelijk begrip hebben van topalbums.

Tenslotte wordt een casestudie gepresenteerd waarin een heterogene groep van experten probeert om een consensus te bereiken over gezamenlijke XBE's. Voor deze casestudie wordt een nieuw consensusproces voorgesteld dat het *"flexible attribute-set consensus reaching"-proces* (FAST-CR) wordt genoemd. De moderator in een FAST-CR-proces geeft aan de deelnemers een collectie van attributen (of kenmerken) die tijdens het evaluatieproces in aanmerking werden genomen door sommige deelnemers, maar onbekend zijn voor andere deelnemers. De moderator kan dan aan een expert vragen om zijn/of haar aandacht te heroriënteren naar voorheen onbekende kenmerken om zo zijn/haar evaluaties te herzien om de mate van consensus over de gezamenlijke XBE's te verhogen. Het FAST-CR-proces is voorgesteld als een belangrijke toepassing die voortvloeide uit de zoektocht naar een oplossing om verschillen in begrip over een evaluatievraag, waarbij XBE's worden gegeven door respondenten met een verschillende achtergrond, adequaat te behandelen.

De bijdragen van dit proefschrift zijn bedoeld om antwoorden te geven op twee vragen die waarschijnlijk zullen rijzen wanneer iemand overweegt om crowdsourcing-diensten te gebruiken: (i) *kan men anonieme respondenten vertrouwen om een taak in zijn/haar plaats uit te voeren?* En (ii) *kan men betrouwbare informatie verkrijgen van onbekende en heterogene bronnen?*

Summary

Experience-based evaluations (XBEs) are appraisals based on what someone has understood or learned about a topic by experience. XBEs can result from an explicit evaluation request or, for example, be a consequence of unsolicited opinions posted on social media. For instance, while the statement "*Airline XYZ* can slightly be considered an *affordable airline* because this airline does not charge any cost for your second luggage" is an XBE given by a person who answered the evaluation request "evaluate to which level *Airline XYZ* can be considered to be an *affordable airline,*" the social media post "*@AirlineXYZ* is an *#AffordableAirline* because of its offers :)" can be deemed to be an XBE that is a consequence of an unrequested opinion about the *Airline XYZ* given by someone who got a nice offer.

Although XBEs can be highly subjective, imprecise, diverse and potentially marked by hesitation, information extracted from them can result in tangible benefits for companies and organizations. Because of this, handling XBEs in a proper way is quite significant in modern information management. Nevertheless, dealing with XBEs can involve several challenging tasks. This is specially the case when potential data quality issues, such as a lack of reliability on XBEs provided by a large and heterogeneous group of (anonymous) sources, need to be handled.

An adequate representation is needed to record and process XBEs given by persons that may have different understandings of the concept under evaluation. Hence, finding a proper representation constitutes a first challenge for dealing with XBEs. For this task, two approaches are proposed in this work: one approach in which nothing about the *context* of an XBE, i.e., the conditions that arise when the appraisal is made, is recorded; and another approach in which hints on the context are recorded. In this regard, existing and novel models and concepts from the area of *computational intelligence*, which aim to find approximate, achievable and robust solutions, are studied and proposed in this dissertation. More specifically, in this work it is studied *if* and *how* concepts like *fuzzy sets*, *intuitionistic fuzzy sets*, *Pythagorean fuzzy sets* and *bipolar satisfaction degrees*, which are all connected to *fuzzy set theory*, can be used for modeling XBEs. Among these concepts, *intuitionistic fuzzy sets* are proven to be adequate and practical for characterizing subjective, imprecise, diverse and marked-by-hesitation XBEs given by respondents having a similar *'mental picture'* or understanding of the concept under evaluation. In

that case, the context of the XBEs do not need to be recorded because it is implicitly assumed that all respondents for their evaluations have focused on the same (or very similar) features of the evaluated objects. In other cases, where XBEs are given by respondents with different understandings, it is not possible to assume that all of them have focused on the same (or very similar) features. So, it is useful to record hints about their appraisals. For such cases, *augmented appraisal degrees* (AADs) and *augmented (Atanassov) intuitionistic fuzzy sets* (AAIFSs), novelties in this work, are proven to be more suitable for representing XBEs.

A second challenge is concerned with the methods and techniques that are necessary to process XBEs. This challenge is connected with the characterization of XBEs because the results obtained after processing two XBEs will depend on how these XBEs are represented. For instance, consider that the statement "*Airline XYZ* can slightly be deemed to be an *affordable airline* because of its recurrent offers" is an XBE given by another person who answered the aforementioned evaluation request. If someone records only the levels of the XBEs resulting from that evaluation request and compare them, both XBEs will match because both XBEs have the same '*slightly affordable airline*' level. On the other hand, if also the stated reasons are recorded, the comparison of these XBEs will produce a different result because the characterizations *'slightly affordable airline'*@{*'no charges on second luggage'*} and *'slightly affordable airline'*@{*'recurrent offers'*} are different. For XBEs that are characterized as elements of an intuitionistic fuzzy set (IFS), existing and novel similarity measures are proposed and used for comparing XBEs within the IFS framework. Furthermore, a novel software package, named *IFSMetrics*, is proposed and described. By means of this software package, such similarity measures can be tested while comparing IFSs that characterize simulated XBEs given by a heterogeneous group of respondents. The results of an empirical study performed with *IFSMetrics* suggest that similarity measures augmented with a kind of *footprint* of the comparison perform reliable comparisons of such simulated XBEs. For XBEs that are characterized as elements of an AAIFS, a novel framework with novel operators and methods is proposed and proved to be suitable for performing reliable comparisons of such XBEs.

A third challenge is related to the *quality of XBEs*. In this work, *high-quality XBEs* are deemed to be XBEs that are fit for use by a requester. This means that *usefulness* and *the extent to which an XBE is fit for use*, are considered to be important aspects of the quality of XBEs. Since the *relevance* of an XBE is tightly linked to its usability, this feature has received special attention in the quality assessment of XBEs. In this regard, the relevance (and, thus, the quality) of the XBEs given by a respondent is considered to be dependent on how aligned the understanding (of the evaluated concept) possessed by this respondent is in relation to the understanding (of the evaluated concept) possessed by the requester. Novel *augmented* similarity measures for the IFS framework, as well as comparison operators and similarity measures for the novel AAIFS framework are proven to be adequate for measuring the level to which XBEs are relevant according to a particular understanding. Hence, both frameworks are deemed to be viable options to measure the perceived quality of XBEs according to a particular understanding.

Contributing to solving these three identified challenges shapes the purpose of this dissertation, which consists in studying *if* and *how* existing or novel concepts and methods in the area of computational intelligence can be used to characterize and process XBEs in such a way that one can adequately handle data quality issues on subjective data provided by a large and heterogeneous group of respondents. An important contribution of this work, obtained while searching for solutions for the identified challenges, is the definition of the novel concept *augmented appraisal degree* (AAD). As indicated above, AADs have been proposed for the characterization of subjective, imprecise, diverse and potentially marked-by-hesitation XBEs in such a way that they straightforwardly lend themselves for computation. This concept along with the definition of other novel related concepts and methods constitute the aforementioned *augmented framework*, which makes it possible to perform a reliable comparison between XBEs given by a heterogeneous group of respondents, as well as to measure the perceived quality of XBEs according to a particular understanding.

IFSMetrics is another important contribution. As was already mentioned, this open-source software package implements an innovative experimental study proposed for testing similarity measures with IFSs characterizing simulated XBEs given by a heterogeneous group of respondents. A significant aspect of *IFSMetrics* is that a researcher or practitioner can use, adapt or extend its code to test other existing or novel similarity measures.

An additional contribution of this dissertation is a novel method, named the *k-well-(un)fitted specimens method*, which has been proposed for computing an approximation of the level to which the contexts of XBEs on social media content are perceived as alike. This method only uses the appraisal levels included in the XBEs of a number of specific social media posts that a requester has considered to be relevant for the concept under study. By means of an empirical study with simulated XBEs, it has been given adequate evidence to consider the k-well-(un)fitted specimens method a suitable tool for handling XBEs for which only appraisal levels are available, as well as for identifying potential differences in the understandings that a heterogeneous group of respondents may have about the concept under evaluation. A novel method, named the *post digest method*, is another contribution in this regard. Posts on social media can be digested by this method to obtain AAIFSs characterizing XBEs. One can use the tools in the augmented framework with the digested AAIFSs to detect opinions (or messages) posted by people sharing a similar understanding about a given fact or topic. The applicability of the post digest method has been illustrated by means of an example in which music album reviews were digested to detect reviewers having a similar understanding of top-rank albums.

Last but not least, a case study in which a heterogeneous group of experts try to reach consensus on *collective XBEs* is presented. For this case study, a novel consensus reaching process, named the *flexible attribute-set consensus reaching process* (FAST-CR), is proposed. The moderator in a FAST-CR process provides the participants with a collection of attributes (or features),

which during the evaluation process were taken into account by some of those participants, but were unobserved by others. The moderator can ask each expert to refocus his/her attention on previously unobserved features and, thus, review his/her evaluations to increase the level of consensus on the collective XBEs. The FAST-CR process has been proposed as a significant application that resulted from the search for a solution to adequately deal with differences in understandings about an evaluation request, in which XBEs are given by respondents with different background.

The contributions of this dissertation are intended to provide answers to two questions that are likely to arise when someone is deciding on the use of crowdsourcing services: (i) *Can someone trust anonymous respondents to perform a job on his/her behalf?* And, (ii) *can someone obtain reliable information from unknown and heterogeneous sources?*

Contents

Abbreviations and Acronyms

A

AAD	Augmented Appraisal Degree
AAF	Augmented Appraisal Function
AAIFS	Augmented (Atanassov) Intuitionistic Fuzzy Set

B

BSD	Bipolar Satisfaction Degree

C

CAF	Connotation Alikeness Factor
CDM	Connotation Differential Marker
CDP	Connotation Differential Print

I

IFS	Intuitionistic Fuzzy Set

P

PFS	Pythagorean Fuzzy Set

S

SVM	Support Vector Machine

X

XBE	Experience-Based Evaluation

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Experience-Based Evaluations

Experience-based evaluations (XBEs), i.e., evaluations resulting from what one has learned or understood about a particular topic by experience, are essential aspects of modern information management. This is specially the case for information handling processes in which someone is trying to find insights in data coming from social media or crowdsourcing. For instance, XBEs of the truthfulness and suitability of postings related to a new governmental regulation can be needed by a social media professional who wants to provide a governmental agency with insights about how to make the regulation more achievable [1, 2].

An XBE can result from a formal and explicit evaluation request. For example, the statement "*tennis* can be strongly considered a *safe sport* because the *players do not have physical contact*" constitutes an XBE given by a respondent who answered the evaluation request "evaluate to which level *tennis* can be deemed a *safe sport*." However, XBEs can also be a consequence of (unsolicited) opinions or judgments. For instance, "*@VeggieChef24* is definitively a *#MustVisitResto* because of its *ratatouille*" can be considered an XBE that result from the opinion about the restaurant "*VeggieChef24*" given by someone who enjoyed its "*ratatouille*."

Since XBEs are linked to the way of thinking that a respondent (or a provider) might have about the topic under analysis, dealing with them can involve several challenging tasks. To get an introduction to these challenges, one can consider the next example in which a group of people is asked to evaluate the level to which a poster is suitable for a new cycle route campaign.

Example 1.1
A group of people is asked to respond to the following request: using a unit interval scale where 1 *represents the highest level of suitability and* 0 *the lowest, evaluate to which degree the poster depicted in Figure 1.1 is* suitable for the new cycle route campaign. *Each response is obtained as shown in Figure 1.2. Notice in the depicted response that a respondent can indicate which feature(s) of the poster got his/her attention during the evaluation process. While two*

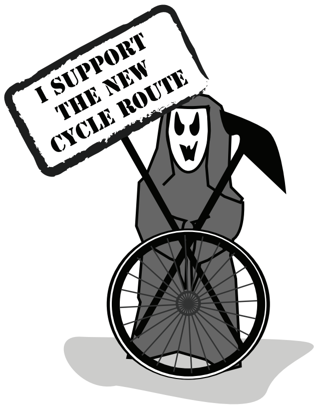


Figure 1.1: Is this poster suitable for the new cycle route campaign?

respondents, say Alice *and* Bob*, assign* 0.45 *and* 0.65 *respectively due to the* text on it*, other respondents, say* Chloe *and* Dexter*, assign* 0.35 *and* 0.67 *in that order because of the* ghost face*.*

A first challenge in this example is concerned with the *characterization of XBEs* since a proper representation is necessary to deal with the XBEs given by Alice, Bob, Chloe and Dexter – e.g., to perform a comparison between these XBEs, a representation of them that lends itself to such a comparison is needed. To address this challenge, one might assume that the stated reasons are irrelevant and, thus, record only the levels expressed in each of the XBEs. If so, one can characterize Alice's, Bob's, Chloe's and Dexter's XBEs as real numbers within the unit interval, i.e., 0.45, 0.65, 0.35 and 0.67 respectively. However, one might also assume that the reasons given by the respondents are necessary for a comparison. If that is the case, one can use structures like 0.45@{*'text on it'*}, 0.65@{*'text on it'*}, 0.35@{*'ghost face'*} and 0.67@{*'ghost face'*} to represent Alice's, Bob's, Chloe's and Dexter's XBEs respectively.

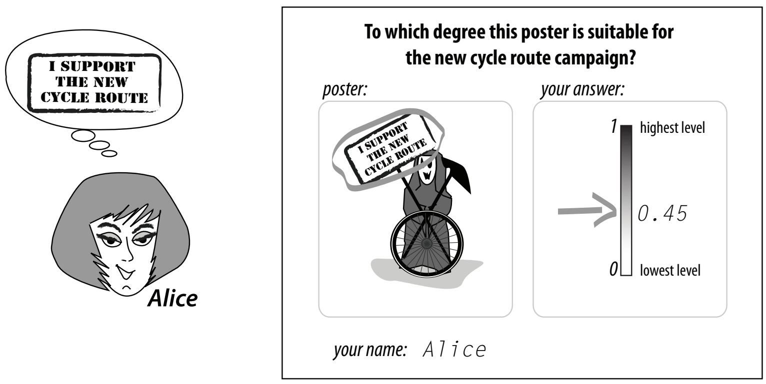


Figure 1.2: Obtaining Alice's experience-based evaluation.

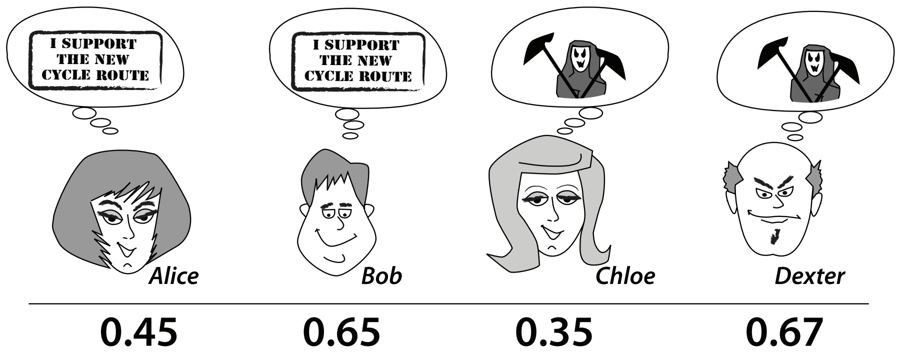


Figure 1.3: Experience-Based Evaluations given by some respondents.

A second challenge is related to the *methods and techniques needed to process such XBEs*. This challenge is connected with the previous since it depends on the characterization of the XBEs. If one wants, e.g., to compare the XBEs and uses one of the aforementioned characterizations, the following results can be obtained. When real numbers are used, Alice's 0.45 will be more similar to the 0.35 from Chloe than the 0.65 from Bob. Also in this case, Dexter's 0.67 and Bob's 0.65 will look very similar to each other. However, when the enhanced, more informative structure is used, Alice's 0.45@{*'text on it'*} will be more similar to Bob's 0.65@{*'text on it'*} than Chloe's 0.35@{*'ghost face'*} and, in addition, Dexter's 0.67@{*'ghost face'*} and Bob's 0.65@{*'text on it'*} will look dissimilar to each other. Notice here how the result of a comparison between XBEs is influenced by their characterization.

A third challenge is about the *quality of the XBEs* since they are inherently subjective and will depend on the standpoints of both a respondent and the *requester*, i.e., the person who asked for the evaluations. If, e.g., the requester in Example 1.1 needs help to evaluate additional posters, one might ask: which respondent(s) should be chosen to perform the evaluations on his/her behalf? In such a situation, one can assume that the respondents with whom the requester shares a similar understanding of *posters suitable for a cycle route campaign* can provide XBEs as good as his/hers. Here, a practical motivation for this dissertation is to address the question: *how to reliably find XBEs from respondents who share a similar understanding with the requester?*

Solving these three challenges shapes the purpose of this PhD study, which will be presented in the next section.

1.2 Purpose of this PhD Study

The research area of *computational intelligence* consists of mathematical models, concepts and methods by which a computer system can show signs of an *intelligent behavior* while performing a complex task [3] – here, by *'intelligent behavior'* is meant the way in which a computer system can use stored data, information or knowledge to deal with new (and sometimes different) situations.

In that regard, the purpose of this dissertation is to study *if* and *how* existing or novel models, concepts and methods in the area of computational intelligence can be used to characterize and process XBEs in such a way that one can identify and manage potential quality issues on subjective data provided by a large and heterogeneous group of (anonymous) people.

To have a better understanding of the purpose of this study, one can consider the structural representation depicted in Figure 1.4. The figure shows three subproblems that result from inherent logical links among the challenges introduced in the previous section. Descriptions of the origin and the importance of solving these subproblems are presented next.

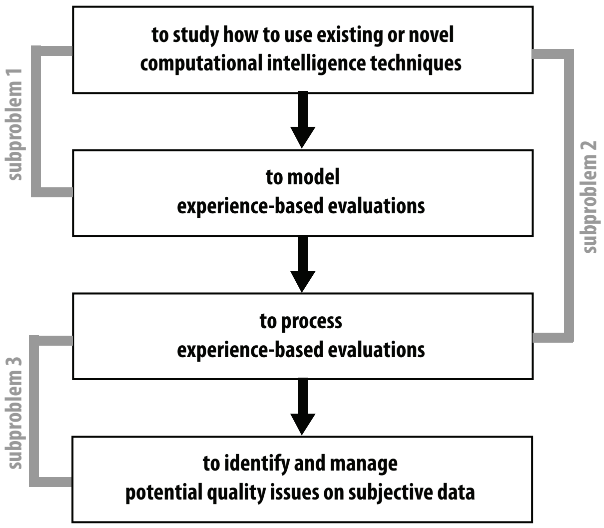


Figure 1.4: A structural representation of the study.

The first subproblem arises from the characterization of XBEs: *how existing or novel computational intelligence models can be used to characterize XBEs in such a way that those XBEs are suitable for computation.* Solving this subproblem is particularly significant since the characterization of XBEs will determine how these XBEs can be stored and used in a forthcoming process – as was mentioned in the previous section, the result of a comparison process between two XBEs can be highly influenced by their characterization.

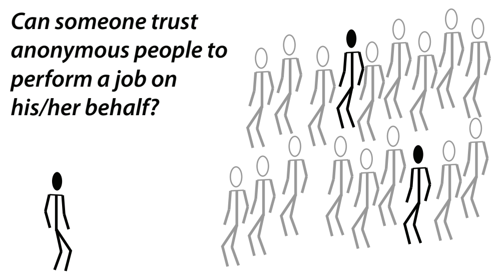
The second subproblem emerges from the need for processing XBEs: *how existing or novel computational intelligence methods can be used to properly process XBEs.* The importance of addressing this subproblem lies in the reliability of the results obtained after processing XBEs – e.g., in the introductory example one can easily argue against the reliability of the result of the similarity comparison between Bob's and Dexter's XBEs when these XBEs are characterized as real numbers. This is even more evident if such a result is used as an input into another process – e.g., one can also argue against including Bob's and Dexter's XBEs in the same group based on the result of the aforementioned similarity comparison.

The third subproblem is evident when it is required to determine the quality of an XBE: *how existing or novel computational intelligence methods or techniques can be used to process XBEs in such a way that one can determine how good (or bad) these XBEs are.* Solving this subproblem is quite important and beneficial to modern information management since the methods or techniques proposed as solutions can be applied to identify and manage potential quality issues on subjective data. For instance, in *crowdsourcing services* in which workers or contributors perform evaluations on behalf of a requester like an editor or social media professional [4, 5], such methods can be applied to identify workers with whom the requester shares a similar understanding of the topic under evaluation. In this case, one can expect that the XBEs provided by these workers, who constitute a *crowdsourced workforce*, will be as good as the XBEs provided by the requester. One can also expect that the quality of the information resulting after processing the XBEs provided by these workers will be better than the quality of the information resulting from XBEs given by people having different understanding. It is worth mentioning that an XBE given by one of these workers is deemed to be the result of a *human-intelligent task*, i.e., a task where *"human intelligence is more efficient or effective than computer analysis"* [4].

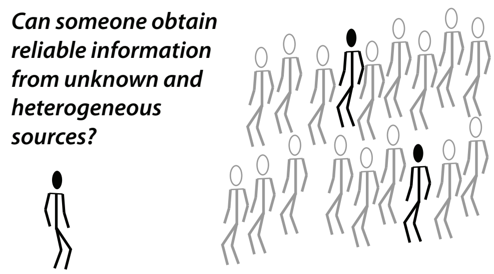
Although finding feasible solutions to the first and second subproblems is a motivation for this work, finding a solution to the third subproblem constitutes the main motivation because of its practical implications. By solving the third subproblem, one can answer two questions that are likely to arise when someone is deciding on the use of crowdsourcing services where human intelligence is needed (see Figure 1.5): (i) *can someone trust anonymous people to perform a job on his/her behalf?* and (ii) *can someone obtain reliable information from unknown and heterogeneous sources?*

1.3 Scope

Studies in the area of computational intelligence aim to find not completely accurate solutions but achievable and robust ones. Such studies are typically based on at least one of the following ideas: (i) nature, i.e., the physical world including plants, animals and other aspects or phenomena, can be a source of inspiration for problem solving; and (ii) imperfect-knowledge can be valuable for people [3]. While examples of studies applying the first idea are *neural networks* (which are inspired by the structure and operation of the brain) and *evolutionary algorithms* (which mimic the principles of biological evolution), examples of studies considering the second idea are *fuzzy systems* (which exploit the fact that human communication and conceptualization are vague) and *probabilistic models* (which are based on probability theory) – see Figure 1.6. In this regard, since XBEs are inherently human statements that can be imperfect (and given that the area of computational intelligence is fairly wide), this



(a)



(b)

Figure 1.5: Practical implications of the main motivation for this dissertation.

dissertation will be mainly focused on studies covering fuzzy systems. More specifically, models, concepts and methods connected with *fuzzy set theory* [6] will be closely studied in this work.

With respect to the kind of data, XBEs that are potentially very subjective, diverse, imprecise and marked by hesitation will be considered for analysis because such XBEs are expected in data provided by a heterogeneous and large group of (usually anonymous) respondents. However, XBEs that are consequence of an explicit evaluation request like the one presented in Example 1.1 will receive more attention in the dissertation because such an evaluation request is usually needed to get *crowdsourced XBEs*, i.e., XBEs carried out by participants in crowdsourcing services where human intelligence is required or expected to perform better than computer analysis. In this regard, we will mainly focus on aspects of crowdsourcing related to activities in which the participants are asked to perform *evaluation tasks* on data that can be hard to process by computers. This means that aspects of crowdsourcing connected to activities such as broadcast search, reporting (or rectifying) problems on data, creative production or development of solutions will be excluded from this dissertation. It is worth mentioning that this delimitation is needed since different (and sometimes conflicting) interpretations of the term '*crowdsourcing*' can be found in the literature – about forty interpretations have been surveyed in [5].

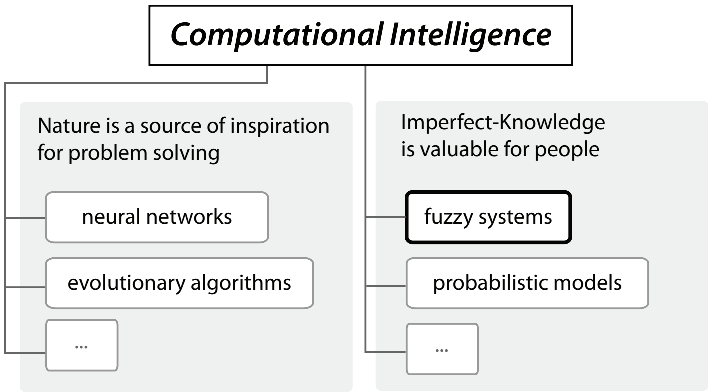


Figure 1.6: Ideas and concepts behind computational intelligence.

Concerning the methods and techniques needed to process XBEs, the dissertation will be mostly focused on *comparison procedures* since they are required to classify, filter or even arrange XBEs. More specifically, comparison procedures aiming to estimate a perceived similarity among XBEs will be studied in this dissertation.

Regarding the quality of XBEs, one can adopt the definition of *high-quality data* proposed by Wang and Strong which states *"high-quality data is data that is fit for use by data consumers"* [7, 8]. By doing so, one can say that *high-quality XBEs* are XBEs that are fit for use by a requester. Hence, one can deem *usefulness*, i.e., the fact that an XBE can be used, and *usability*, i.e., the level to which an XBE is fit to be used, to be important aspects of the quality of XBEs. Furthermore, one might be tempted to use attributes like *accuracy*, *objectivity* and *completeness* as important aspects of the quality of XBEs – a set of these attributes representing a single aspect of data quality is also known as a *data quality dimension* in the data quality literature [7, 8, 9, 10]. However, such attributes might not be possessed by XBEs since, as said above, XBEs can be very subjective, diverse, imprecise and marked by hesitation. For this reason, only attributes like *believability*, *relevance* or *value-added* that might be related to the quality of an XBE could be considered for its assessment. In this work special attention will be given to the *relevance* of an XBE since this attribute is strongly linked to the *perceived quality* of an XBE, which can be used by a requester to assess the quality of crowdsourced XBEs – recall that handling properly crowdsourced XBEs is a practical motivation for this work. Such an approach, in which only a few of the attributes connected to the quality of data are taken into account, has also been followed in [11, 12]. Figure 1.7 shows some of the attributes traditionally associated to the quality of data: while the attributes that might by considered to assess the quality of XBEs are represented with bold lines, the attributes that could not be considered are represented with gray lines.

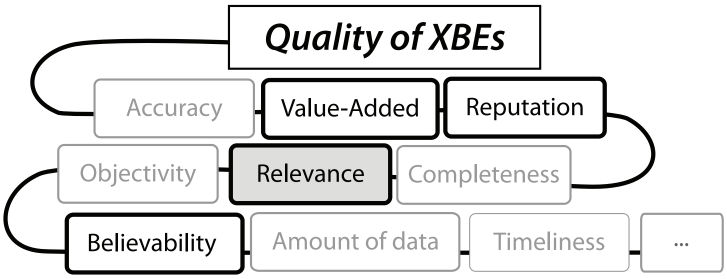


Figure 1.7: Attributes that might be linked to the quality of XBEs.

1.4 Related Work

1.4.1 About the characterization of XBEs

A group of respondents can be deemed to be *homogeneous* when its members have the same (or very similar) understanding of the concept under study, which can, e.g., be obtained by instructing or training them. In this regard, although not being explicitly mentioned as such, several frameworks that might be applicable for the characterization of XBEs given by a *homogeneous group of people* can be found in the literature.

One of such frameworks is the *fuzzy set theory* [6]. In that framework, an XBE can be seen as the result of judging a perceived belongingness of a subject to a given concept and, thus, it can be characterized by means of a *membership grade*, which is a real number in the unit interval $[0, 1]$ that indicates the extent to which such a subject belongs to the concept. In this case, the *degree-of-similarity* semantic interpretation of a membership grade presented in [13], in which a membership grade denotes the level to which a subject is similar to the prototype of the concept under study, can be applied. For instance, if one is asked to assess the level to which each hotel in a collection, say $X = \{$*'Hotel 1'*, *'Hotel 2'*, *'Hotel 3'*$\}$, is deemed to be part of a collection of *high standard hotels*, say A, one can establish a membership grade for each hotel, say $\mu_A(\textit{'Hotel 1'}) = 0.2$, $\mu_A(\textit{'Hotel 2'}) = 0.7$ and $\mu_A(\textit{'Hotel 3'}) = 0.6$ respectively, according to the perceived similarity between each hotel and what one understands as a high standard hotel. Together these membership grades constitute a *fuzzy set* of high standard hotels. Such a fuzzy set can mathematically be denoted by

$$A = \{\langle x, \mu_A(x)\rangle \,|\, (x \in X) \wedge (0 < \mu_A(x) \leq 1)\}, \tag{1.1}$$

where x characterizes a hotel in X, A represents the (fuzzy) subset of X under consideration, e.g., the subset of high standard hotels, and $\mu_A(x)$ denotes the level to which x is member of A.

It is worth mentioning that, in the fuzzy set theory framework, an XBE can also be seen as the result of the evaluation of a proposition p having a canonical form '*x IS A*', which means "the value of x is compatible with the definition of A" [14]. Thus, e.g., $\mu_A(\text{'}\textit{Hotel 2}\text{'}) = 0.7$ can be seen as an XBE that results after evaluating the proposition p : "'*Hotel 2' IS A*", which means "*Hotel 2* is an instance of (the fuzzy set of) high standard hotels." Notice here that the result of the evaluation of p is expressed as a *matter of degree*, i.e., the result is not limited to both a full agreement and a full disagreement, but all the values in between. This is an important aspect of modeling an XBE as an element of a fuzzy set, since a respondent will be able to evaluate a proposition even if he/she is not fully convinced about its truthfulness.

In situations where a respondent hesitates about the level to which an object belongs to a fuzzy set A, his/her XBE can be better described in the framework of *intuitionistic fuzzy sets*, IFS for short [15, 16]. In the IFS framework, the XBE of an object, say x, can be characterized as an *IFS element*, say $\langle x, \mu_A(x), \nu_A(x)\rangle$, in which the components $\mu_A(x)$ and $\nu_A(x)$ represent respectively the levels of *membership* and *nonmembership* of x to IFS A. Hence, a collection of XBEs can be represented as an IFS A defined by

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | \, (x \in X) \wedge (0 \leq \mu_A(x) + \nu_A(x) \leq 1)\}, \qquad (1.2)$$

where X symbolizes a collection containing the objects that are being evaluated, x characterizes an object in X, A is a subset of X related to the concept under evaluation and, finally, $\mu_A(x)$ and $\nu_A(x)$ denote the membership and nonmembership degrees of x in A respectively – here, the condition $0 \leq \mu_A(x) + \nu_A(x) \leq 1$, also known as *consistency condition*, reflects the (implicit) assumption that $\mu_A(x)$ and $\nu_A(x)$ depend on each other. A geometrical interpretation of an IFS element is shown in Figure 1.8. Notice that the consistency condition is represented by the longest side of the (right) triangle since no point (i.e., no IFS element) can be placed outside this triangle.

As an example of the characterization of XBEs in the framework of IFSs, consider the next IFS which represents the XBEs given by a respondent about the hotels in the above-mentioned example:

$$A = \{\langle \text{'}\textit{Hotel 1}\text{'}, 0.2, 0.5\rangle, \langle \text{'}\textit{Hotel 2}\text{'}, 0.7, 0.1\rangle, \langle \text{'}\textit{Hotel 3}\text{'}, 0.6, 0.4\rangle\}.$$

Notice here that the XBE of *'Hotel 1'* given by this respondent indicates the level to which he/she considers this hotel as a member of (the collection of) high standard hotels, as well as the level to which the hotel is not part of that group. Notice also that this respondent hesitates about these levels, which follows from the expression $\mu_A(\text{'}\textit{Hotel 1}\text{'}) + \nu_A(\text{'}\textit{Hotel 1}\text{'}) = 0.7 < 1$. This kind of *hesitation margin* was also proposed by Atanassov in [15, 16] and is defined by

$$h_A(x) = 1 - \mu_A(x) - \nu_A(x). \qquad (1.3)$$

Enabling the characterization of the hesitation that a person might have while providing his/her XBEs can be deemed to be an advantage of the IFSs in comparison to the traditional (or standard) fuzzy sets.

In [17], Atanassov proposed another geometric interpretation of an IFS element in which the consistency condition is denoted by $0 \leq (\mu_A(x))^2 +$

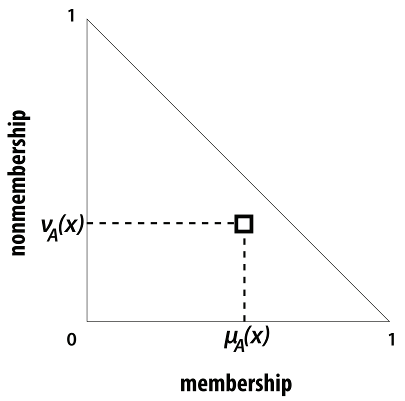


Figure 1.8: A geometrical interpretation of an IFS element.

$(\nu_A(x))^2 \leq 1$. As shown in Figure 1.9, this interpretation increases the space of possible values that can be used when an XBE is characterized as an IFS element. Using this interpretation, Atanassov introduced and defined the concept of an *IFS of second type*, or IFS-2T for short, as follows [17, 18]:

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle \,|\, (x \in X) \wedge \big(0 \leq (\mu_A(x))^2 + (\nu_A(x))^2 \leq 1\big)\}. \quad (1.4)$$

The hesitation margin in this case is defined by

$$h_A(x) = \sqrt{1 - (\mu_A(x))^2 - (\nu_A(x))^2}. \quad (1.5)$$

As an example of the characterization of XBEs as an IFS-2T, consider the next IFS-2T which represents the XBEs given by another respondent:

$$A = \{\langle \textit{`Hotel 1'}, 0.1, 0.6\rangle, \langle \textit{`Hotel 2'}, 0.5, 0.8\rangle, \langle \textit{`Hotel 3'}, 0.7, 0.3\rangle\}.$$

In this case, the increased space of values of an IFS-2T is used for denoting the XBE of *'Hotel 2'*. Notice that the consistency condition for an IFS-2T holds since $0 \leq (0.5)^2 + (0.8)^2 \leq 1$.

Another potential framework in which XBEs might be characterized is the framework of *Pythagorean fuzzy sets*, PFS for short, proposed by Yager in [19, 20]. In this framework, the XBE of an object, say x, can be characterized as a *Pythagorean membership grade*, say $\langle x, r_A(x), d_A(x)\rangle$, in which the components $r_A(x)$ and $d_A(x)$ represent respectively the *strength* and the *direction* of the commitment (membership or nonmembership) that a respondent might have while judging x to be part of the PFS A. A geometrical interpretation of a Pythagorean membership grade is depicted in Figure 1.10. Notice that, while a value of $d_A(x)$ close to 1 means "support for membership of x in A," a value

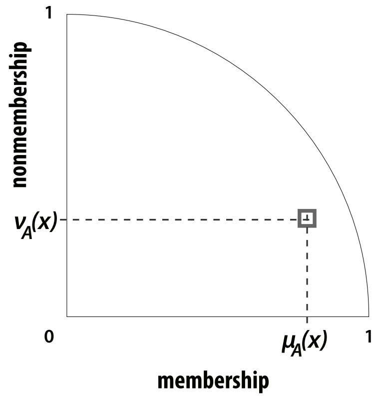


Figure 1.9: A geometrical interpretation of an IFS element of second type.

close to 0 means "support against membership of x in A" [19, 20]. Notice also that a *neutral position* about the membership of x in A is denoted by $d_A(x) = 0.5$. In this regard, a collection of XBEs can be characterized as a PFS A defined by

$$A = \{\langle x, r_A(x), d_A(x)\rangle \,|\, (x \in X) \wedge (r_A(x) \in [0,1]) \wedge (d_A(x) \in [0,1])\}. \quad (1.6)$$

Here, in a similar way to the definition of an IFS, X denotes a collection containing the objects that are being evaluated, x characterizes an object in X and A is the PFS under evaluation.

According to their definitions, an IFS and a PFS can mathematically be related, as shown in Figure 1.11, by means of the equations

$$\mu_A(x) = r_A(x)\cos(\theta(x)) \quad (1.7)$$

and

$$\nu_A(x) = r_A(x)\sin(\theta(x)), \quad (1.8)$$

where

$$\theta(x) = (1 - d_A(x))\frac{\pi}{2}. \quad (1.9)$$

For this reason, a PFS can be deemed to be an IFS-2T [18] even though the motivations behind these concepts are different.

As an example of the characterization of XBEs in the framework of PFSs, consider the following PFS characterizing the XBEs of the hotels in the above example

$$A = \{\langle \textit{'Hotel 1'}, 0.54, 0.24\rangle, \langle \textit{'Hotel 2'}, 0.71, 0.91\rangle, \langle \textit{'Hotel 3'}, 0.72, 0.63\rangle\}.$$

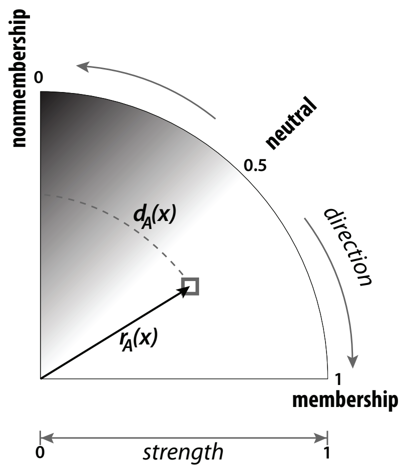


Figure 1.10: A geometrical interpretation of a Pythagorean membership grade.

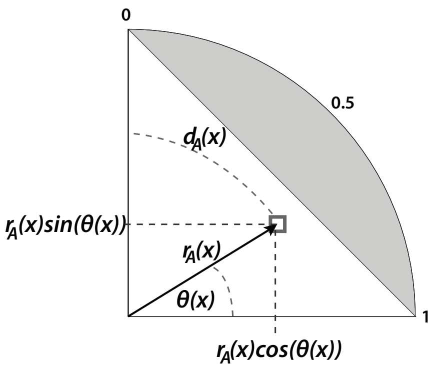


Figure 1.11: Pythagorean membership grade vs IFS element.

In this case, while the XBE of *'Hotel 2'* denotes a rather considerable support for the membership in the collection of high standard hotels because r_A(*'Hotel 2'*) $= 0.71$ and d_A(*'Hotel 2'*) $= 0.91$, the XBE of *'Hotel 1'* suggests a medium support against the membership of this hotel because r_A(*'Hotel 1'*) $=$ 0.54 and d_A(*'Hotel 1'*) $= 0.24$.

As can be noticed, PFSs increase the space of possibles values as IFSs of second type do. However, factors such as the added complexity of the computation needed to process these variants of IFSs can favor the selection of IFSs defined by (1.2) to characterize XBEs.

An option that increases the space of values without adding extraordinary complexity is given in the framework of *Bipolar Satisfaction Degrees*, BSDs for short [21, 22]. In that framework, the XBE of an object x satisfying a proposition p :'x *IS* A' can be modeled as a pair $\langle s_A(x), d_A(x)\rangle$ such that $s_A(x)$ and $d_A(x)$ belong to the unit interval $[0,1]$ and represent, respectively, the level to which x *satisfies* and the level to which x *dissatisfies* p. Although from a semantical point of view the definition of a BSD is closely related to the definition of an IFS element, a BSD does not impose the constraint $0 \leq s_A(x) + d_A(x) \leq 1$ because these values are deemed to be independent of each other and, thus, it is allowed that $s_A(x) + d_A(x) > 1$. A geometrical interpretation of a BSD is depicted in Figure 1.12. This figure shows how a BSD can be specified: when the BSD lies on the light gray zone, i.e., when $s_A(x) + d_A(x) < 1$, the BSD is *underspecified*; when the BSD lies on the bold line, i.e., when $s_A(x) + d_A(x) = 1$, the BSD is *fully specified*; and when the BSD lies on the dark gray zone, i.e., when $s_A(x) + d_A(x) > 1$, the BSD is *overspecified*. The figure also shows three special cases: *full dissatisfaction*, i.e., when $s_A(x) = 0$ and $d_A(x) = 1$, which is denoted by $\circ$; *full satisfaction*, i.e., when $s_A(x) = 1$ and $d_A(x) = 0$, which is represented by $\diamond$; and a *neutral position*, i.e., when $s_A(x) = d_A(x)$, which is illustrated by the dotted line.

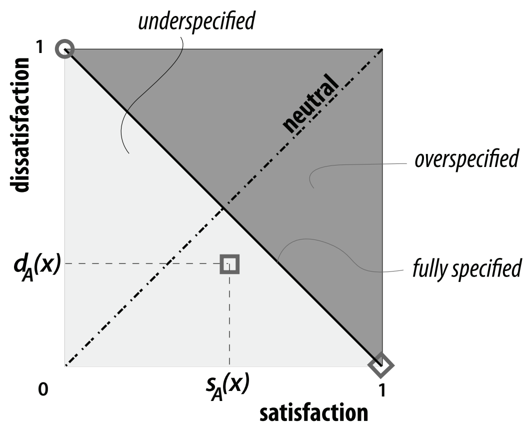


Figure 1.12: A geometrical interpretation of a BSD.

As an example of the characterization of XBEs in the framework of BSDs, consider that the BSD $\langle 0.75, 0.47\rangle$ characterizes an XBE of *'Hotel 3'* given by a respondent in the aforementioned example. In this case, this BSD is deemed to be overspecified since s_A(*'Hotel 3'*) $+ d_A$(*'Hotel 3'*) > 1. This suggests that the respondent sees something that makes *'Hotel 3'* a member of the collection of high standard hotels, but, at the same time, the respondent sees something else that excludes this hotel of that collection.

While all the characterizations of XBEs presented so far make it possible to record the level to which a respondent considers a proposition is true, none of these characterizations enables the recording of the reasons that the respondent might give to justify that level. Hence, they might not be applicable to model XBEs given by a *heterogeneous group* of people since one could not assume that all the respondents focus on the same (or very similar) aspects of a subject during its evaluation. This is one of the reasons why the characterization of such XBEs constitutes a challenge in this dissertation.

1.4.2 About the comparison of XBEs

After modeling a collection of XBEs by a fuzzy set, an IFS or a PFS, one can theoretically use any of the existing similarity measures designed for comparing such sets to perform a comparison between XBEs.

In the literature, two kinds of approaches in the formulation of such similarity measures can be identified: *symmetric* and *directional* approaches. In the former, also known as *metric distance approaches*, the similarity between two fuzzy sets, IFSs or PFSs, say A and B, is considered to be symmetrical. This means that, if $S(A, B)$ denotes the level to which A is similar to B, in a symmetric approach the expression $S(A, B) = S(B, A)$ always hold. In contrast, in a directional approach, the similarity between A and B is deemed to be the result of judging a statement having the form '*A is like B*,' where a *subject* A and a *referent* B can be identified. Furthermore, in a directional approach is considered that judging the statement '*A is like B*' is not always equivalent to judging '*B is like A*' [23]. For instance, in the statement '*Person 1* is like *Barack Obama*' is suggested that the characteristics of '*Person 1*' are being compared with the political qualities of (the referent) '*Barack Obama*.' However, in the statement '*Barack Obama* is like *Person 1*,' the characteristics of '*Barack Obama*' might be compared with unknown characteristics of '*Person 1*.' In other words, in a directional approach $S(A, B)$ might not be equal to $S(B, A)$.

Studies that follow a symmetric approach usually define a similarity measure as a function of a *metric distance*, say $d(A, B)$, in such a way that

$$S(A, B) = 1 - d(A, B). \tag{1.10}$$

For instance, consider two fuzzy sets, say A and B, in a finite referential collection X. If $d(A, B)$ is a function that computes the *normalized Euclidean distance* [24] between A and B by means of the equation

$$d(A, B) = \sqrt{\frac{1}{n} \sum_{x \in X} |\mu_A(x) - \mu_B(x)|^2}, \tag{1.11}$$

where n represents the number of elements in X, a *normalized Euclidean similarity measure* [25] that obtains the similarity between A and B will be defined by

$$S_E(A,B) = 1 - \sqrt{\frac{1}{n}\sum_{x\in X}|\mu_A(x) - \mu_B(x)|^2}. \qquad (1.12)$$

In a similar way, if $d(A,B)$ is a function that computes the normalized *Euclidean* distance between *IFSs* A and B [16] by means of the equation

$$d(A,B) = \sqrt{\frac{1}{2n}\sum_{x\in X}(|\mu_A(x) - \mu_B(x)|^2 + |\nu_A(x) - \nu_B(x)|^2)}, \qquad (1.13)$$

a *normalized Euclidean similarity measure* that obtains the similarity between IFSs A and B [25] will be defined by

$$S_E(A,B) = 1 - \sqrt{\frac{1}{2n}\sum_{x\in X}(|\mu_A(x) - \mu_B(x)|^2 + |\nu_A(x) - \nu_B(x)|^2)}. \qquad (1.14)$$

In [26], Szmidt and Kacprzyk studied several similarity measures that compare IFSs by means on a distance function. In that work, difficulties on the design of similarity measures applying this approach have been identified. To overcome those difficulties, the inclusion of the *complement of an IFS element*, i.e., $\langle x, \mu_A(x), \nu_A(x)\rangle^C = \langle x, \nu_A(x), \mu_A(x)\rangle$, into the definition of a similarity measure has been proposed. An extension of this study was presented in [27].

With respect to the framework of PFSs, although a deeper study about similarity measures in that framework is to the best of our knowledge absent in the literature, a distance measure designed to compare Pythagorean fuzzy sets can be found in [28].

Regarding similarity measures following a directional approach, in [29] a definition of a similarity measure that compares traditional fuzzy sets with an extension of the *feature contrast model* [23] was proposed. In the feature contrast model, the similarity between two objects, say o_1 and o_2, is mathematically described by the expression

$$S(o_1, o_2) = \lambda_1 \cdot f(O_1 \cap O_2) - \lambda_2 \cdot f(O_1 - O_2) - \lambda_3 \cdot f(O_2 - O_1), \qquad (1.15)$$

where O_1 and O_2 represent the collection of features identified in o_1 and o_2 respectively, λ_1, λ_2 and λ_3 are non-negative numbers, and f is a non-negative measure of the contribution of the common features or the features that belong exclusively to either o_1 or o_2. In the extension proposed in [29], while the objects o_1 and o_2 constitute traditional fuzzy sets, f is a function that computes the intersection and the difference between such fuzzy sets.

In [23], Tversky also proposed the *ratio model*, in which the similarity is described by the expression

$$S(o_1, o_2) = \frac{f(O_1 \cap O_2)}{f(O_1 \cap O_2) + \lambda_2 \cdot f(O_1 - O_2) + \lambda_3 \cdot f(O_2 - O_1)}. \qquad (1.16)$$

As shown in [30], this model can be seen as a generalization of other similarity measures. For instance, *Jaccard similarity measure* [31] can be obtained by setting $\lambda_2 = 1$ and $\lambda_3 = 1$ in (1.16). Similarly, *Dice similarity measure* [32] can be obtained by setting $\lambda_2 = 0.5$ and $\lambda_3 = 0.5$ in (1.16).

Although the aforementioned studies include the theoretical analysis about the applicability of the similarity measures, they do not provide evidence on their performance while comparing XBEs. Because such an analysis might provide insights about the reliability of a similarity measure while comparing XBEs, it is deemed to be an important part of this dissertation.

1.4.3 About the quality of XBEs

Studies that might be applied to assess the quality of XBEs given by a large number of respondents can be found in the literature. In most of these studies, an approach in which the quality of subjective data is evaluated through *crowdsourcing services*, i.e., services where *workers* or *contributors* perform a job on behalf of a *requester* [4, 5], is taken. Therein, the research efforts are usually oriented to detect and recruit qualified workers.

To identify qualified workers, those studies typically apply methods based on a *gold standard* collection, which contains questions with correct answers that allow a requester to rate the reliability of each worker [33, 34]. In this regard, various strategies to build such collections exist.

One of such strategies consists in building a gold standard collection with questions and answers given by experts. This technique was applied in [35] to assess the quality of non-experts annotations in contrast to annotations given by experts. Since this strategy depends on the time (and cost) of experts, it could result in gold collections having a limited number of elements. To avoid this limitation, a strategy in which a gold collection is compiled using answers resulting from an agreement among workers has been presented in [36]. A variant of this strategy that includes the accuracy of a worker and the number of answers provided by him/her has been presented in [37].

Another strategy, in which questions with correct answers are generated on the basis of questions with known answers, was proposed in [38]. In that work, the authors introduce a particular type of errors in the well-known answers to generate new answers. The idea is that a reliable worker will be able to detect such errors and choose the proper answer.

To compute the reliability of a worker, the studies based on gold standard collections in most of the cases assume that the answers are precise and unaffected by any difference in understandings that the workers might have. With that assumption, one can compute a *gold score* for a worker by means of the equation

$$s_{gold} = \frac{g}{G}, \tag{1.17}$$

where g and G represent the number of correctly answered questions and the size of the gold collection respectively. The idea behind this equation is that a reliable worker will be someone who is able to answer correctly all or a large number of the questions included in the gold collection. A potential weakness of the aforementioned assumption is the introduction of strict restrictions when workers are trying to express their XBEs.

Other studies propose to keep track of the characteristics and skills of the workers in order to choose among them the more reliable ones. In this regard, attributes like expertise or scientific credit can be asked to a worker to assess his/her quality [39]. Such attributes can also be acquired in a dynamic way to build an ontology of skills, i.e., a collection of concepts and categories that show the skills of a worker and the relation between them [40].

In the case of vague and heterogeneous knowledge, an ontology of skills can be characterized by means of *fuzzy ontologies*, i.e., a fuzzy set of concepts and categories that show the skills of a worker and the (fuzzy) relation between [41, 42]. By doing so, one can perform a comparison between the skills of a worker and the skills of a requester to detect a potential alignment between them [43, 44].

In contrast to the aforementioned studies, in this dissertation the *context of an XBE*, i.e., the conditions that arise when the evaluation is made, is taking into account to assess its quality. The rationale behind this approach is that when a respondent agrees with a requester on the reasons given during an evaluation, the requester will find the evaluation more reliable.

1.5 Research Questions

The next questions have been phrased according to the challenges and the purpose of the study presented in the above sections:

Q1. How to characterize subjective, imprecise and potentially marked-by-hesitation XBEs in such a way that they are suitable for computation?

Q2. How to perform a reliable comparison between XBEs given by a heterogeneous group of respondents?

Q3. How to measure the perceived quality of XBEs according to a particular understanding?

Q4. How to identify XBEs given by (anonymous) respondents with whom a requester shares a similar understanding of the topic under analysis?

Q5. How to detect and manage automatically any difference in understanding of a concept behind an evaluation request, in which the answers could be given by respondents with different background?

The contributions that result after answering these questions will be presented in the next part.

1.6 Main Contributions

A main contribution of this dissertation is a novel concept named *augmented appraisal degree*, AAD for short, [45], which has been proposed while answering Research Question Q1 to characterize not just the level but also the reasons expressed in an XBE. Along with this concept, an *augmented framework* constituted by other novel related concepts that make it possible to model (i.e., Research Question Q1) and compare (i.e., Research Questions Q2 and Q3) XBEs given by a heterogeneous group of respondents has been proposed. As an example, the definition of an *augmented appraisal function* has been proposed to represent a collection of AADs provided by a particular respondent.

An innovative experimental test proposed while answering Research Questions Q2 and Q3 with XBEs characterized as IFSs is another important contribution of this work. This test has been implemented in an open-source software package named *IFSMetrics* [46]. Through this package, one or more (configurations of) similarity measures can be tested with a big number of IFSs characterizing XBEs that result from different learning scenarios. Reports resulting from the package show that only a few of the similarity measures existing in the literature reflect properly a perceived similarity when IFSs resulting from opposite scenarios are compared to each other. Since IFSs have been increasingly applied to deal with problems in topics like pattern recognition, *IFSMetrics* can be deemed to be an important tool to improve the reliability of similarity comparisons applied in such topics.

An additional contribution of this dissertation is a novel method proposed while answering Research Questions Q4 and Q5. The method, named the *k-well-(un)fitted specimens method*, aims to obtain an approximation of the level to which the contexts of XBEs on social media content are perceived as alike. To do so, the method relies only on the appraisal levels included in the XBEs of a specific number of social media posts that a requester deems to be representative of the concept under study. Since by means of the k-well-(un)fitted specimens method a requester can determine how good (or bad) for him/her the XBEs given by a respondent are, it can be considered to be an appropriate instrument to identify a right collaborator for a particular task. For instance, after computing the approximations of the contexts of XBEs given by a group of anonymous respondents, a requester can empower the respondents with whom he/she shares a similar understanding to act on his/her behalf.

Last but not least, a novel consensus reaching process, in which the options considered within a decision-making process are evaluated by people having different expertise, has been proposed as an unintended but significant consequence while answering Research Question Q5. An interesting aspect of this process, which is called *flexible attribute-set consensus reaching process* or *FAST-CR* for short, is that attributes (or features) of the options initially unobserved by some participants can be made available to the others. Thus, a participant can refocus his/her attention on such previously unobserved features and, thus, review his/her evaluations to increase the level of consensus throughout the (consensus reaching) process.

1.7 Outline

To present the aforementioned contributions, the dissertation has been structured as follows:

- Chapter 2 presents an interpretation on how the knowledge acquired by a respondent can be reflected in his/her XBEs. Such interpretation facilitates the understanding of the aspects that might have an influence on the context of XBEs. Hence, it is deemed to be needed to address the challenges described in the previous sections.
- Chapters 3, 4 and 5 explore the implications of characterizing XBEs as IFSs. Chapter 3 first describes how a collection of XBEs can be characterized as an IFS and then presents a novel approach to compare any two of them. Chapter 4 presents a novel procedure by which similarity measures designed to compare IFSs are tested in order to determine how suitable these measures are for comparing XBE sets. Chapter 5 describes *IFSMetrics*, an open-source software package that implements the aforementioned test procedure.
- Chapters 6, 7, 8 and 9 explore the implications of characterizing XBEs as AADs. Chapter 6 describes the concept of AADs and how it can be applied to handle XBEs. Chapter 7 presents a novel method to identify and handle context in *plain* XBEs, i.e., XBEs in which only the levels of appraisal are given. Chapter 8 covers a novel technique for handling XBEs when there is no explicit evaluation request. Chapter 9 explores an application of AADs to reach consensus in decision making problems involving a heterogeneous group of experts [47].
- Chapter 10 concludes this dissertation and presents some suggestions about further research.

The above structure is depicted in Figure 1.13:

1.8 Publications

Papers produced during this PhD study have been published or submitted for publication in scientific journals, as well as presented at several international conferences. The following is a full list of these papers:

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

1. Loor Marcelo and De Tré Guy. On the Need for Augmented Appraisal Degrees to Handle Experience-Based Evaluations. *Applied Soft Computing*, 54 (2017): 284-295.

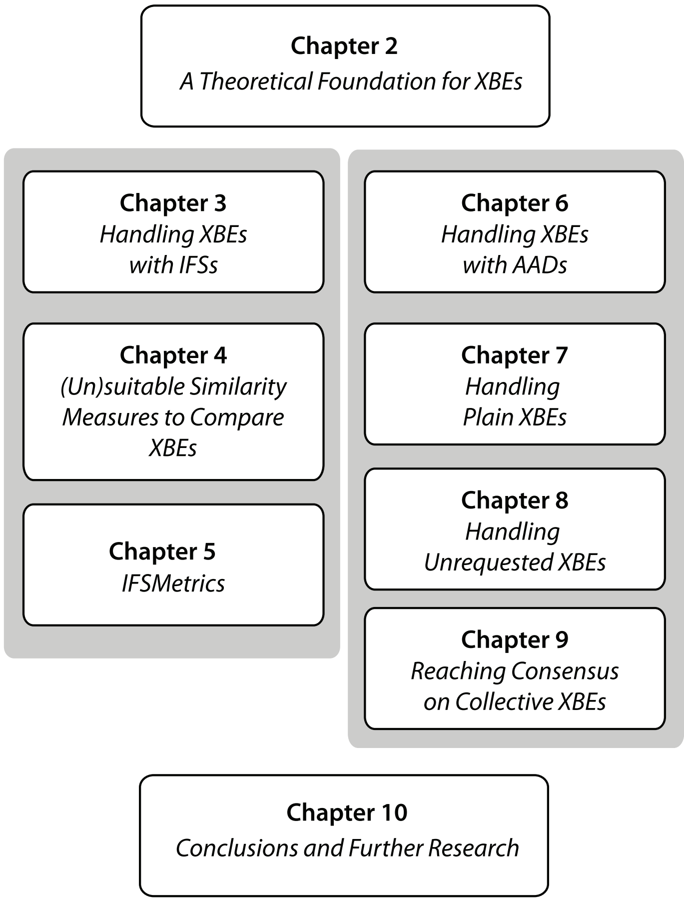


Figure 1.13: Dissertation outline.

2. Loor Marcelo and De Tré Guy. Identifying and Properly Handling Context in Crowdsourcing. *Applied Soft Computing, Under Review.*

3. Loor Marcelo and De Tré Guy. Enabling Augmented (Fuzzy) Computation in Social Media Mining. *Fuzzy Sets and Systems, Under Review.*

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

1. Loor Marcelo and De Tré Guy. In a Quest for Suitable Similarity Measures to Compare Experience-Based Evaluations. In Juan Julián Merelo, Agostinho Rosa, José M. Cadenas, António Dourado Correia, Kurosh Madani, António Ruano, and Joaquim Filipe, editors, Computational Intelligence: International Joint Conference, IJCCI 2015 Lisbon, Portugal, November 12-14, 2015, Revised Selected Papers, volume 669 of *Studies in Computational Intelligence*, pages 291–314. Springer International Publishing, 2017.

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

1. Loor Marcelo and De Tré Guy. Connotation-Differential Prints - Comparing What Is Connoted Through (Fuzzy) Evaluations. In *Proceedings of the International Conference on Fuzzy Computation Theory and Applications - Volume 1: FCTA, (IJCCI 2014)*, 127-136. Rome, Italy, 2014.

2. Loor Marcelo and De Tré Guy. Choosing Suitable Similarity Measures to Compare Intuitionistic Fuzzy Sets that Represent Experience-Based Evaluation Sets. In *Proceedings of the 7th International Joint Conference on Computational Intelligence*, 57-68. Lisbon, Portugal, 2015.

3. Loor Marcelo and De Tré Guy. An Open-Source Software Package to Assess Similarity Measures that Compare Intuitionistic Fuzzy Sets. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–6. Naples, Italy, 2017.

4. Loor Marcelo, Tapia-Rosero Ana, and De Tré Guy. Refocusing Attention on Unobserved Attributes to Reach Consensus in Decision Making Problems Involving a Heterogeneous Group of Experts. In *Proceedings of: EUSFLAT- 2017 – The 10th Conference of the European Society for Fuzzy Logic and Technology, September 11-15, 2017, Warsaw, Poland IWIFSGN'2017 – The Sixteenth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets, September 13-15, 2017, Warsaw, Poland, Volume 2*, pages 405–416, 2018. Springer International Publishing.

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

1. Loor Marcelo and De Tré Guy. Vector Based Similarity Measure for Intuitionistic Fuzzy Sets. In *Modern approaches in fuzzy sets, intuitionistic fuzzy sets, generalized nets and related topics. Volume I: Foundations* edited by Krassimir Atanassov, Michal Baczynski, Jozef Drewniak, Janusz Kacprzyk, Maciej Krawczak, Eulalia Szmidt, Maciej Wygralak and Slawomir Zadrozny, 125-142, SRI-PAS, 2014.

1.9 Awards/Acknowledgments

The following awards/acknowledgments have been received during this PhD study:

- *Best Student Paper Award*[1] at the 7th International Joint Conference on Computational Intelligence, for the paper *Choosing Suitable Similarity Measures to Compare Intuitionistic Fuzzy Sets that Represent Experience-Based Evaluation Sets.* Lisbon, Portugal, 2015.

- *Invited to the Fuzz-IEEE 2017 Doctoral Consortium*[2] at the Fuzz-IEEE 2017 Conference, for the paper *An Open-Source Software Package to Assess Similarity Measures that Compare Intuitionistic Fuzzy Sets.* Naples, Italy, 2017.

[1] http://www.fcta.ijcci.org/PreviousAwards.aspx
[2] http://fuzzieee2017.org/doctoral.html

References

[1] Andrea L. Kavanaugh, Edward A. Fox, Steven D. Sheetz, Seungwon Yang, Lin Tzy Li, Donald J. Shoemaker, Apostol Natsev, and Lexing Xie. *Social media use by government: From the routine to the critical.* Government Information Quarterly, 29(4):480 – 491, 2012.

[2] Ines Mergel. *Social media adoption and resulting tactics in the U.S. federal government.* Government Information Quarterly, 30(2):123 – 130, 2013.

[3] Rudolf Kruse, Christian Borgelt, Christian Braune, Sanaz Mostaghim, and Matthias Steinbrecher. *Computational intelligence: a methodological introduction.* Springer, 2016.

[4] Daren C. Brabham. *Crowdsourcing.* Mit Press, 2013.

[5] Enrique Estellés-Arolas and Fernando González-Ladrón-de Guevara. *Towards an integrated crowdsourcing definition.* Journal of Information science, 38(2):189–200, 2012.

[6] Lotfi A. Zadeh. *Fuzzy sets.* Information and control, 8(3):338–353, 1965.

[7] Richard Y. Wang and Diane M. Strong. *Beyond accuracy: What data quality means to data consumers.* Journal of management information systems, 12(4):5–33, 1996.

[8] Diane M. Strong, Yang W. Lee, and Richard Y. Wang. *Data Quality in Context.* Commun. ACM, 40(5):103–110, May 1997.

[9] Leo L. Pipino, Yang W. Lee, and Richard Y. Wang. *Data Quality Assessment.* Commun. ACM, 45(4):211–218, April 2002.

[10] Carlo Batini, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. *Methodologies for Data Quality Assessment and Improvement.* ACM Comput. Surv., 41(3):16:1–16:52, July 2009.

[11] Sandra de F. Mendes Sampaio, Chao Dong, and Pedro Sampaio. *DQ2S: A framework for data quality-aware information management.* Expert Systems with Applications, 42(21):8304 – 8326, 2015.

[12] Guy De Tré, Robin De Mol, and Antoon Bronselaer. *Handling veracity in multi-criteria decision-making: A multi-dimensional approach.* Information Sciences, 2017.

[13] Didier Dubois, Walenty Ostasiewicz, and Henri Prade. *Fuzzy Sets: History and Basic Notions.* In Didier Dubois and Henri Prade, editors, Fundamentals of Fuzzy Sets, volume 7 of *The Handbooks of Fuzzy Sets Series*, pages 21–124. Springer US, 2000.

[14] Lotfi A. Zadeh. *Is there a need for fuzzy logic?* Information Sciences, 178(13):2751 – 2779, 2008.

[15] Krassimir T. Atanassov. *Intuitionistic fuzzy sets.* Fuzzy sets and Systems, 20(1):87–96, 1986.

[16] Krassimir T. Atanassov. *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[17] Krassimir T. Atanassov. *Geometrical interpretations of the elements of the intuitionistic fuzzy objects.* Preprint IM-MFAIS, 1–89. Sofia (1989). Reprinted in: Int. J. Bioautomation, 20(S1):S27–S42, 2016.

[18] Krassimir T. Atanassov and Peter Vassilev. *On the Intuitionistic Fuzzy Sets of n-th Type*, pages 265–274. Springer International Publishing, 2018.

[19] Ronald R. Yager and Ali M. Abbasov. *Pythagorean Membership Grades, Complex Numbers, and Decision Making.* International Journal of Intelligent Systems, 28(5):436–452, 2013.

[20] Ronald R. Yager. *Pythagorean fuzzy subsets.* In IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint, pages 57–61, June 2013.

[21] Tom Matthé and Guy De Tré. *Bipolar query satisfaction using satisfaction and dissatisfaction degrees: bipolar satisfaction degrees.* In Proceedings of the 2009 ACM symposium on Applied Computing, pages 1699–1703. ACM, 2009.

[22] Tom Matthé, Guy De Tré, Sławomir Zadrożny, Janusz Kacprzyk, and Antoon Bronselaer. *Bipolar database querying using bipolar satisfaction degrees.* International Journal of Intelligent Systems, 26(10):890–910, 2011.

[23] Amos Tversky. *Features of similarity.* Psychological review, 84(4):327, 1977.

[24] Janusz Kacprzyk. *Multistage Fuzzy Control: A Prescriptive Approach.* John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1997.

[25] Eulalia Szmidt and Janusz Kacprzyk. *Entropy for intuitionistic fuzzy sets.* Fuzzy Sets and Systems, 118(3):467 – 477, 2001.

[26] Eulalia Szmidt and Janusz Kacprzyk. *Geometric similarity measures for the intuitionistic fuzzy sets.* In 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13), pages 840–847. Atlantis Press, 2013.

[27] Eulalia Szmidt. *Similarity Measures between Intuitionistic Fuzzy Sets.* In Distances and Similarities in Intuitionistic Fuzzy Sets, volume 307 of *Studies in Fuzziness and Soft Computing*, pages 87–129. Springer International Publishing, 2014.

[28] Xiaolu Zhang and Zeshui Xu. *Extension of TOPSIS to Multiple Criteria Decision Making with Pythagorean Fuzzy Sets.* International Journal of Intelligent Systems, 29(12):1061–1078, 2014.

[29] Simone Santini and Ramesh Jain. *Similarity measures.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(9):871–883, Sep 1999.

[30] Marie-Jeanne Lesot, Maria Rifqi, and Hamid Benhadda. *Similarity measures for binary and numerical data: a survey.* International Journal of Knowledge Engineering and Soft Data Paradigms, 1(1):63–84, 2008.

[31] Paul Jaccard. *Nouvelles Recherches Sur la Distribution Florale.* 44:223–70, 01 1908.

[32] Lee R. Dice. *Measures of the amount of ecologic association between species.* Ecology, 26(3):297–302, 1945.

[33] Petros Venetis and Hector Garcia-Molina. *Quality control for comparison microtasks.* In Proceedings of the first international workshop on crowdsourcing and data mining, pages 15–21. ACM, 2012.

[34] Vikas C. Raykar and Shipeng Yu. *Eliminating spammers and ranking annotators for crowdsourced labeling tasks.* Journal of Machine Learning Research, 13(Feb):491–518, 2012.

[35] Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Y. Ng. *Cheap and Fast—but is It Good?: Evaluating Non-expert Annotations for Natural Language Tasks.* In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 254–263. Association for Computational Linguistics, 2008.

[36] Chris Callison-Burch. *Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk.* In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1, EMNLP '09, pages 286–295, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[37] Sujoy Chatterjee and Malay Bhattacharyya. *Judgment analysis of crowdsourced opinions using biclustering.* Information Sciences, 375:138–154, 2017.

[38] David Oleson, Alexander Sorokin, Greg P Laughlin, Vaughn Hester, John Le, and Lukas Biewald. *Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing.* Human computation, 11(11), 2011.

[39] Gloria Bordogna, Paola Carrara, Laura Criscuolo, Monica Pepe, and Anna Rampini. *A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science.* Information Sciences, 258:312–327, February 2014.

[40] Kinda El Maarry, Wolf-Tilo Balke, Hyunsouk Cho, Seung-won Hwang, and Yukino Baba. *Skill Ontology-Based Model for Quality Assurance in Crowdsourcing*, chapter 1, pages 376–387. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[41] Juan A. Morente-Molinera, Ignacio J. Pérez, Raquel Ureña, and Enrique Herrera-Viedma. *Creating knowledge databases for storing and sharing people knowledge automatically using group decision making and fuzzy ontologies.* Information Sciences, 328:418–434, 2016.

[42] Natalia Díaz Rodríguez, Manuel P Cuéllar, Johan Lilius, and Miguel Delgado Calvo-Flores. *A fuzzy ontology for semantic modelling and recognition of human behaviour.* Knowledge-Based Systems, 66:46–60, 2014.

[43] Marielba Zacarias, Helena Sofia Pinto, R. Magalhães, and J. Tribolet. *A "context-aware"and agent-centric perspective for the alignment between individuals and organizations.* Information Systems, 35(4):441–466, 2010.

[44] Cristina Sarasua, Elena Simperl, and Natalya F. Noy. *CrowdMap: Crowdsourcing Ontology Alignment with Microtasks.* In The Semantic Web – ISWC 2012: 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part I, pages 525–541. Springer Berlin Heidelberg, 2012.

[45] Marcelo Loor and Guy De Tré. *On the Need for Augmented Appraisal Degrees to Handle Experience-Based Evaluations.* Applied Soft Computing, pages 284 – 295, 2017.

[46] Marcelo Loor and Guy De Tré. *An Open-Source Software Package to Assess Similarity Measures that Compare Intuitionistic Fuzzy Sets.* In 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pages 1–6, July 2017.

[47] Marcelo Loor, Ana Tapia-Rosero, and Guy De Tré. *Refocusing Attention on Unobserved Attributes to Reach Consensus in Decision Making Problems Involving a Heterogeneous Group of Experts.* In Janusz Kacprzyk, Eulalia Szmidt, Slawomir Zadrożny, K. T. Atanassov, and Maciej Krawczak, editors, Advances in Fuzzy Logic and Technology 2017: Proceedings of: EUSFLAT- 2017 – The 10th Conference of the European Society for Fuzzy Logic and Technology, September 11-15, 2017, Warsaw, Poland, pages 405–416. Springer International Publishing, 2018.

Chapter 2

A Theoretical Foundation for Experience-Based Evaluations

Abstract

As was presented in Chapter 1, an experience-based evaluation (XBE) is a judgment that results from what one has learned or understood about a particular topic by experience. To assess how good (or bad) such an XBE is, one can take into account its *context*, i.e., the conditions that arise when the XBE is carried out. In this regard, understanding the aspects that might have an influence on the context of XBEs can provide insights about the selection of proper mechanisms to perform such an assessment. Aiming to facilitate such understanding, in this chapter we describe an interpretation on (i) *how a person can experience (or learn) a concept* and (ii) *how the knowledge acquired by this person can be reflected in his/her XBEs.* The interpretation is based on an intuition suggesting that one can learn about a concept by studying objects that satisfy or dissatisfy a criterion related to the concept. An example of this interpretation shows that, although the same learning rules are followed, different experiences of a topic might conduct to different understandings of it, which is then reflected in different XBEs.

This chapter contains parts included in the following papers:

- Marcelo Loor and Guy De Tré. *Vector Based Similarity Measure for Intuitionistic Fuzzy Sets. Modern approaches in fuzzy sets, intuitionistic fuzzy sets, generalized nets and related topics. Volume I: Foundations* edited by K. Atanassov, M. Baczynski, J. Drewniak, J. Kacprzyk, M. Krawczak, E. Szmidt, M. Wygralak and S. Zadrozny, 125-142, SRI-PAS, 2014.
- Marcelo Loor and Guy De Tré. *Identifying and Properly Handling Context in Crowdsourcing. Applied Soft Computing, Under Review.*
- Marcelo Loor and Guy De Tré. *Enabling Augmented (Fuzzy) Computation in Social Media Mining. Fuzzy Sets and Systems, Under Review.*

2.1 Introduction

Imagine a situation in which two cousins, say *Pia* and *Rod*, were asked to evaluate to which degree a cookie can be considered to be similar to Grandma's cookies. While Pia focused on aspects like the *square shape* and the *square hole* that made her think the cookie is a Grandma's, Rod focused his attention on the *linear icing* that made him think the cookie was not baked by Grandma (see Figure 2.1). When asked about her cousin's evaluation, Pia said that Rod seems to be joking because his judgment does not reflect what a Grandma's cookie is.

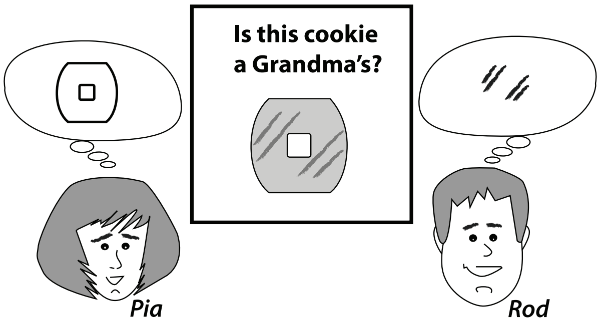


Figure 2.1: Two cousins focusing on different features while evaluating a cookie.

One can notice in the above situation that Pia and Rod focus their attention on different aspects while performing their evaluations. One can also notice that these aspects allow them to judge if the cookie is similar to a Grandma's cookie. In this case, one can ask: *what makes these aspects arise when an XBE is carried out?*

To provide insights on that regard, in this chapter we present an interpretation on how Pia and Rod could learn about what a Grandma's cookie is and how the acquired knowledge is then reflected in their XBEs. To do so, in the next part we describe a (learning) process by which a concept (e.g., *'Grandma's cookies'*) is learned by studying the features (e.g., *'square hole'* or *'linear icing'*) of some objects (e.g., *'cookies'*) that satisfy or dissatisfy an evaluation criterion related to the concept (e.g., to consider a cookie to be a Grandma's cookie, the cookie has to *"look like a Grandma's"*). Then, we explain an evaluation process that uses the acquired knowledge to assess the level to which other objects satisfy or dissatisfy that evaluation criterion.

2.2 An Experience-Based Learning Process

The question raised in this section is about learning: *How can a person experience (or learn) a concept?* To find an answer to this question, one can consider the following intuition: *to learn about a concept, one can study some* *objects that satisfy or dissatisfy an evaluation criterion related to the concept.* This means that one can learn about a concept by looking into the features of some objects that favor or disfavor the fulfillment of an evaluation criterion related to the concept – as will be shown in Section 2.2, after learning about a concept one can, e.g., recognize a pattern for this concept and use it to categorize other (new) objects [1]. To illustrate such an experience-based learning process, let us consider a running example where the individual experiences that the aforementioned cousins have about a Grandma's cookie conduct to different understandings of what a Grandma's cookie is.

Example 2.1
Two cousins learning about what a Grandma's cookie is. *What can Pia and Rod learn about a Grandma's cookie after trying some cookies?*

1. *Pia has received from her dad five cookies (cookies* $1, 2, 3, 4$ *and* 5 *in Figure 2.2), some of them made by Grandma (cookies* $1, 2, 4$ *and* 5*). Pia's dad told her that a Grandma's cookie usually has two distinguishable features: a square shape and a square hole.*

2. *Rod is a cousin of Pia. He has received from his mom, who is sister of Pia's dad, also five cookies (cookies* $1, 2, 3, 4$ *and* 7 *in Figure 2.2), three of them made by Grandma (cookies* $1, 2$ *and* 7*). Rod's mom told him that Grandma prefers making cookies without icing.*

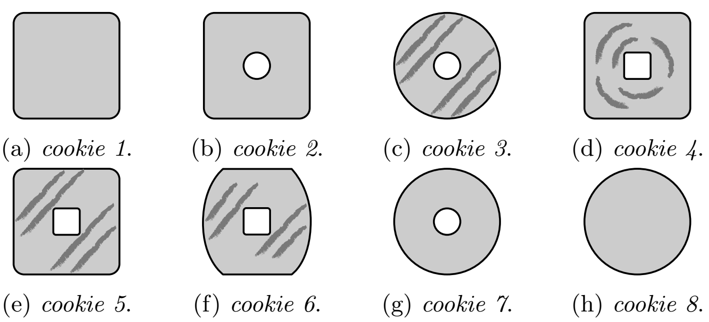


Figure 2.2: Grandma's cookies learning example.

2.2.1 Influence of a feature

Relying on the above intuition, when the fulfillment of a criterion is appraised on an object, some of its *features*, i.e., some of its distinctive attributes, will be more influential (favoring or disfavoring the fulfillment of the criterion) than others. For instance, when the fulfillment of the criterion *"look like a Grandma's cookie"* is appraised by Pia on *cookie 5* (see Figure 2.3), she can find the *square shape* and the *square hole* of this cookie more influential than its *linear icing* because of her dad's suggestion – this kind of suggestion is known as *prior knowledge* in the pattern recognition literature [1].

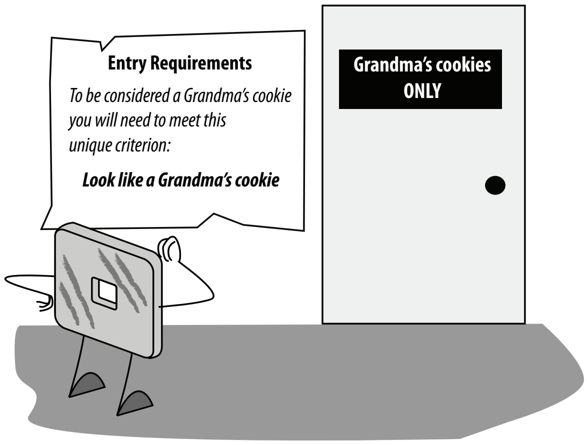


Figure 2.3: Criterion for being considered a Grandma's cookie.

In that regard, one can assume that the influence of a feature on the appraisal of a criterion depends on both its *weight* and its *direction*: while the weight denotes *how influential (or important) a feature is in relation to the others*, the direction denotes *whether the feature's influence is 'in favor of' or 'in opposition to' the criterion.* For instance, Pia can consider the feature *square shape* to be twice more influential (or important) than the feature *linear icing* in order to label or not *cookie 5* as a cookie satisfying (or fulfilling) the criterion *"look like a Grandma's cookie"* [2]. If so, she can assign 2 and 1 as the weights of the *square shape* and the *linear icing* respectively to denote their relative importance – the relative importance of a feature can be used, e.g., to suggest which features should be taken into account during a forthcoming learning process. Furthermore, Pia can consider that the *square shape* is in favor of such a label and, thus, she can say that influence of the *square shape* is 'in favor of' considering *cookie 5* as a cookie satisfying the aforementioned criterion.

One can also assume that the level to which an object satisfies (or dissatisfies) an evaluation criterion will depend on the combined effect of (two or more of) the object's features. For instance, Pia can anticipate that *cookie* 7 will not satisfy the criterion *"look like a Grandma's cookie"* since both the *round hole* and the *round shape* of this cookie are in opposition to this criterion according to what her dad told her about the features of a Grandma's cookie.

2.2.2 A representational model

To denote properly the influence of a feature on the appraisal of an evaluation criterion related to a concept, one can use the following *feature-influence* representational model.

Consider a concept A, i.e., consider that A represents an idea or a group of objects having particular shared aspects – e.g., consider that A represents

Grandma's cookies. Consider also a criterion $\mathcal{C}$ having a form like "be compatible with the way in which the concept A is perceived" – e.g., consider the criterion "be compatible with the way in which a Grandma's cookie is perceived." Let $X = \{x_1, \cdots, x_n\}$ be a collection of objects, where each $x_i \in X$ has a collection of features $\mathcal{F}_i$ – e.g., let $X = \{\textit{cookie 1}, \textit{cookie 2}, \textit{cookie 3}, \textit{cookie 4}, \textit{cookie 5}\}$ be the collection of cookies received by Pia where, e.g., *cookie* 2 has a collection of features $\mathcal{F}_2 = \{\textit{round-hole}, \textit{square-shape}\}$. Assume $\mathcal{F} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_n$. Consider an m-dimensional feature space $\mathcal{M}$ in which each dimension corresponds to a feature $f_j \in \mathcal{F}$, i.e., m corresponds to the number of features in $\mathcal{F}$. In this context, the *feature-influence* representational model is conceived as follows:

- A vector $\mathbf{f}_{i,j} = \beta_{i,j}\hat{\mathbf{f}}_j$ in $\mathcal{M}$ represents the *overall influence* of f_j when the fulfillment of $\mathcal{C}$ is appraised on x_i, i.e, when the proposition "x_i satisfies $\mathcal{C}$" is appraised. Herein $\beta_{i,j}$ is a real number that represents the *overall weight* (or importance) of f_j among the features in x_i (e.g., if f_1 is deemed to be three times more important than f_2, the weights 3 and 1 can be assigned to $\beta_{i,1}$ and $\beta_{i,2}$ respectively), and $\hat{\mathbf{f}}_j$ is a unit vector that represents the dimension corresponding to f_j in $\mathcal{M}$. When the appraisal is not influenced by f_j, $\beta_{i,j} = 0$ holds. Consequently, the *resulting overall influence* of the features of x_i on such appraisal is represented by a vector $\mathbf{x}_i$ such that $\mathbf{x}_i = \mathbf{f}_{i,1} + \cdots + \mathbf{f}_{i,m} = \beta_{i,1}\hat{\mathbf{f}}_1 + \cdots + \beta_{i,m}\hat{\mathbf{f}}_m$, i.e., the resulting overall influence corresponds to the vector sum of all the overall influences of the features in $\mathcal{F}$. For instance, in Figure 2.4 the resulting overall influence of the features of x_i on the appraisal of the proposition "x_i satisfies $\mathcal{C}$" is represented by $\mathbf{x}_i = \beta_{i,1}\hat{\mathbf{f}}_1 + \beta_{i,2}\hat{\mathbf{f}}_2$, which is a vector in a two-dimensional feature space whose dimensions correspond to the features f_1 and f_2 respectively. Semantically, the overall influence of a feature, say f_j, represents the *prior knowledge* that someone might have about the contribution of f_j on the appraisal of the proposition "x_i satisfies $\mathcal{C}$."

- A particular understanding (or knowledge) about the concept A, say K_A, is characterized by both $\hat{\mathbf{u}}_A = \omega_1\hat{\mathbf{f}}_1 + \cdots + \omega_m\hat{\mathbf{f}}_m$ and t_A, and represented as a line in $\mathcal{M}$, where $\hat{\mathbf{u}}_A$ is a unit vector called *directional vector* and t_A is a point on the line called *threshold point*. The direction of $\hat{\mathbf{u}}_A$ is defined towards a place in $\mathcal{M}$ where the fulfillment of $\mathcal{C}$ is favored, while the opposite direction is defined towards a place where the fulfillment of $\mathcal{C}$ is disfavored. With regard to t_A, its location identifies a point (on the line K_A) where the fulfillment of $\mathcal{C}$ is neither favored nor disfavored. For example, the lines $K_{A@P}$ and $K_{A@Q}$, which represent the understandings about concept A possessed by persons P and Q respectively, are shown in Figure 2.5. Notice in this figure the relative alignment between the understandings possessed by P and Q.

- A vector $\mathbf{f}_{i,j_A} = \beta_{i,j_A}\hat{\mathbf{u}}_A$ represents the *specific influence* of $f_j \in \mathcal{F}_i$ on x_i when the proposition "x_i satisfies $\mathcal{C}$" is appraised, and corresponds to the *vector projection* of (its overall-influence vector) $\mathbf{f}_{i,j}$ on $\hat{\mathbf{u}}_A$. Herein β_{i,j_A} is a real number that represents the *specific weight* of this feature

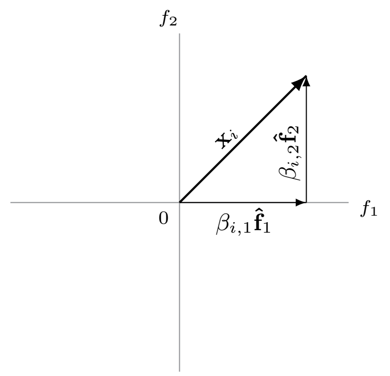


Figure 2.4: Resulting overall influence of the features f_1 and f_2 for the appraisal of the criterion $\mathcal{C}$ on x_i .

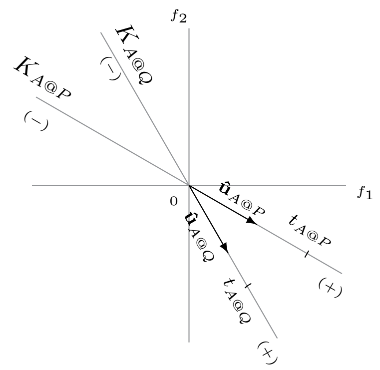


Figure 2.5: Characterization of two particular understandings (or knowledge) about the concept A.

during such appraisal. The direction of $\mathbf{f}_{i,j_A}$ determines whether f_j is *"in favor of"* or *"in opposition to"* the fulfillment of $\mathcal{C}$: when $\hat{\mathbf{u}}_A$ and $\mathbf{f}_{i,j_A}$ have the same direction, it is the former case; and when $\hat{\mathbf{u}}_A$ and $\mathbf{f}_{i,j_A}$ are opposite to each other, it is the latter case. For instance, in Figure 2.6 the *specific influences* of the features f_1 and f_2 on the appraisal of the proposition "x_i satisfies $\mathcal{C}$" are depicted as the *vector projections* of (the overall-influence vectors) $\mathbf{f}_{i,1}$ and $\mathbf{f}_{i,2}$ respectively. Notice in this figure that while f_1 is *in favor of* the fulfillment of $\mathcal{C}$, f_2 is *in opposition to* it. One can say that, while vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ respectively denote the *intended* or *believed* influence of features f_1 and f_2 on the appraisal of the proposition "x_i satisfies $\mathcal{C}$", vectors $\mathbf{f}_{i,1}$ and $\mathbf{f}_{i,2}$ denote the *actual* influence of these features.

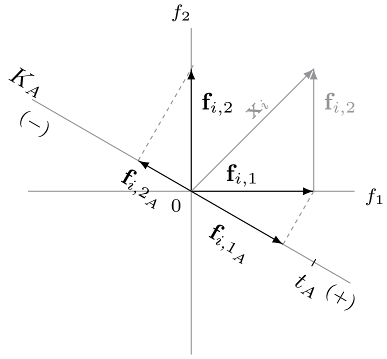


Figure 2.6: Specific influence of $f_j \in \mathcal{F}_i$ on x_i when the proposition "x_i satisfies $\mathcal{C}$" is appraised.

- The *resulting specific influence* of the features of the object x_i on the judgment of the proposition "x_i satisfies $\mathcal{C}$", say $\mathbf{x}_{iA}$, is the vector sum of all the specific influences of these features, i.e., $\mathbf{x}_{iA} = \mathbf{f}_{i,1_A} + \cdots + \mathbf{f}_{i,m_A}$ or $\mathbf{x}_{iA} = (\beta_{i,1_A} + \cdots + \beta_{i,m_A})\hat{\mathbf{u}}_A$. The level to which x_i satisfies (or dissatisfies) $\mathcal{C}$ is given by the magnitude of vector
$$\mathbf{l}_{iA} = \mathbf{x}_{iA} - t_A\hat{\mathbf{u}}_A. \tag{2.1}$$
When the direction of $\mathbf{l}_{iA}$ is opposite to the direction of $\hat{\mathbf{u}}_A$, x_i dissatisfies $\mathcal{C}$ at a level given by $\|\mathbf{l}_{iA}\|$, where $\|\mathbf{l}_{iA}\|$ is the magnitude of $\mathbf{l}_{iA}$ (i.e., $\sqrt{\mathbf{l}_{iA} \cdot \mathbf{l}_{iA}}$). Opposed to that, when the directions of $\mathbf{l}_{iA}$ and $\hat{\mathbf{u}}_A$ are the same, x_i satisfies $\mathcal{C}$ at a level given by $\|\mathbf{l}_{iA}\|$. Since the level to which x_i satisfies (or dissatisfies) $\mathcal{C}$ can be seen as a matter of degree, x_i can be labeled as an object that, e.g., *'partially satisfies (or dissatisfies) $\mathcal{C}$'* depending on this level. For instance, in Figure 2.7 the resulting influence of the features of x_i, namely f_1 and f_2, on the appraisal of $\mathcal{C}$ is depicted as $\mathbf{x}_{iA}$. Notice here that, although the resulting influence is in favor of (the fulfillment of) $\mathcal{C}$, it does not overcome the threshold t_A and, thus, "x_i dissatisfies $\mathcal{C}$ at a level $\|\mathbf{l}_{iA}\|$" is obtained as a result. In this example, x_i could be labeled as an *object that dissatisfies $\mathcal{C}$*.

As can be noticed, not only the influence of the features over the appraisal of an evaluation criterion on an object but also the level of this appraisal have geometrical interpretations in the feature-influence representational model. Hence, one can use this model to visualize the features that might have an influence on the context of an XBE resulting from the appraisal of, e.g., the proposition '*cookie* 5 satisfies the criterion "be compatible with the way in which a Grandma's cookie is perceived."' It could also be noticed that, by means of this model, one can visualize how an object satisfies (or dissatisfies) an evaluation criterion related to one or more concepts. Later on, these aspects will be used to illustrate different understandings of a given concept.

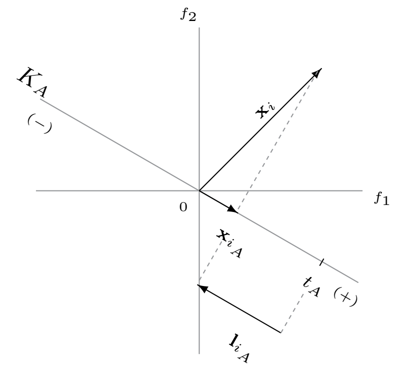


Figure 2.7: Resulting specific influence of the features of the object x_i on the judgment of the proposition "x_i satisfies $\mathcal{C}$".

2.2.3 Gathering knowledge about concept A

As previously stated, an XBE mainly depends on the experience (or knowledge) acquired by a respondent about the concept under evaluation, say A. Since the respondent can obtain that experience in different ways, in this part we describe a learning method that mimics a learning behavior in which one could learn about (the concept) A by studying (the features of) objects that satisfy or dissatisfy an evaluation criterion related to A.

The learning method relies on two inputs, say X_0 and Y_0, to get a model of the knowledge about A, say K_A. While X_0 contains objects that either satisfy or dissatisfy a criterion related to A, say $\mathcal{C}$, Y_0 consists of labels that indicate which objects in X_0 satisfy and which ones dissatisfy $\mathcal{C}$. Since both X_0 and Y_0 can be used to learn about A, they are usually deemed to be constituents of a *training data set*. For instance, in Figure 2.8 $X_{0@Pia}$ and $Y_{0@Pia}$ constitute the inputs for the learning method followed by Pia while learning about A.

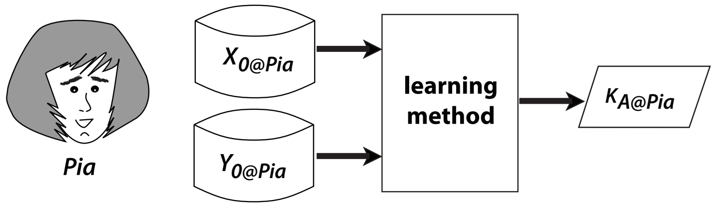


Figure 2.8: Pia's learning process.

The main steps of the learning method are the following:

1. Identify the (collection of) features that constitute each object $x_i \in X_0$ and put all of them into a collection $\mathcal{F}$.
2. Assign an overall weight $\beta_{i,j}$ to each feature $f_j \in \mathcal{F}$ according to its overall influence on the appraisal of (the fulfillment of) $\mathcal{C}$ on each $x_i \in X_0$.
3. Compute the constituents of K_A, i.e., $\hat{\mathbf{u}}_A = \omega_1\hat{\mathbf{f}}_1 + \cdots + \omega_m\hat{\mathbf{f}}_m$ and t_A, in such a way that (i) the correspondence between each $x_i \in X_0$ (not) fulfilling $\mathcal{C}$ and the resulting specific influence of its features are preserved, and (ii) both the aggregate of the specific influences of the features in favor of $\mathcal{C}$ and the aggregate of the specific influences of the features in opposition to $\mathcal{C}$ are maximized.
4. Return both $\hat{\mathbf{u}}_A$ and t_A that maximize both aggregates.

In the first step, the objects' features that will be taken into account during the learning process need to be identified. For example, to learn about a Grandma's cookie, Pia can study the cookies that satisfy or dissatisfy the evaluation criterion "the cookie is compatible with the way in which A is perceived," where A represents (the concept) *Grandma's cookies*. To do so, she can use a training data set consisting of the cookies given by her dad along with the labels indicating whether or not a cookie is made by Grandma. Using the above notation, Pia can represent such cookies and labels as $X_{0@Pia} = \{x_1, x_2, x_3, x_4, x_5\}$ and $Y_{0@Pia} = \{y_1, y_2, y_3, y_4, y_5\}$ respectively, where, e.g., x_5 corresponds to *cookie* 5 and y_5 corresponds to a label like the number 1 indicating that x_5 is made by Grandma. In such a case, the following features can respectively be detected in cookies x_1, x_2, x_3, x_4, x_5:

$$\begin{aligned}
\mathcal{F}_1 &= \{\textit{square-shape}, \textit{no-icing}\},\\
\mathcal{F}_2 &= \{\textit{square-shape}, \textit{no-icing}, \textit{round-hole}\},\\
\mathcal{F}_3 &= \{\textit{round-shape}, \textit{linear-icing}, \textit{round-hole}\},\\
\mathcal{F}_4 &= \{\textit{square-shape}, \textit{curved-icing}, \textit{square-hole}\}, \text{ and}\\
\mathcal{F}_5 &= \{\textit{square-shape}, \textit{linear-icing}, \textit{square-hole}\}.
\end{aligned}$$

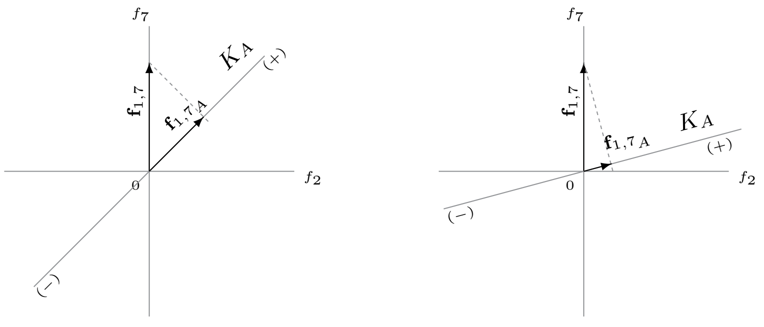
Hence, the collection of features $\mathcal{F}$ that can be taken into account during the Pia's learning process will result from $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \cup \mathcal{F}_4 \cup \mathcal{F}_5$. This collection is listed in the first column of Table 2.1.

In the second step, an overall weight for each of the already identified features has to be assigned according to its relative overall importance (or influence) on the appraisal of (the criterion) $\mathcal{C}$. For example, since Pia's dad told her that a square shape and a square hole are distinguishable features in a Grandma's cookie, Pia can consider that the features related to the shape of a cookie and the shape of a hole on it are twice more important than the others. If so, she can assign the overall weights of each feature $f_j \in \mathcal{F}$ as shown in the third column of Table 2.1. Notice in this table that, since 1 has been assigned as an overall weight to a "non distinguishable" feature like *linear-icing*, the weight of a distinguishable feature like *round-shape* has been set to 2.

Table 2.1: Collection of features identified by Pia.

Pia		
feature	f_j	overall weight
linear-icing	f_1	1
no-icing	f_2	1
round-hole	f_3	2
curved-icing	f_4	1
round-shape	f_5	2
square-hole	f_6	2
square-shape	f_7	2

The third step could become more complex. To illustrate how it works, let us start by describing how the specific influence of a feature can be modified by adjusting the constituents of K_A, i.e., $\hat{\mathbf{u}}_A$ and t_A.



(a) (b)

Figure 2.9: Modifying the specific influence of a feature.

Consider the representation of the influence of the feature *'square-shape'* shown in Figure 2.9a. In this representation, while the overall influence of *'square-shape'* on *cookie* 1 is depicted by $\mathbf{f}_{1,7} = \beta_{1,7}\hat{\mathbf{f}}_7$, the specific influence of this feature on the appraisal of *cookie* 1 satisfying $\mathcal{C}$ is illustrated by $\mathbf{f}_{1,7_A} = \beta_{1,7_A}\hat{\mathbf{u}}_A$. Notice that the magnitude of $\mathbf{f}_{1,7_A}$, i.e., $||\mathbf{f}_{1,7_A}|| = |\beta_{1,7_A}|$, decreases when the line depicting K_A is turned clockwise as shown in Figure 2.9b.

Now consider the representation of the influence of the two features identified in *cookie* 1 (see Figure 2.10). In this representation, while the resulting *overall* influence of *cookie* 1's features, i.e., *'square-shape'* ($\mathbf{f}_{1,7}$) and *'no-icing* ($\mathbf{f}_{1,2}$), is depicted by $\mathbf{x}_1 = \mathbf{f}_{1,7} + \mathbf{f}_{1,2}$, the resulting *specific* influence these features on the appraisal of *cookie* 1 satisfying $\mathcal{C}$ is illustrated by $\mathbf{x}_{1_A}$ (see Figure 2.10a). In a similar way to what happened with the magnitude of $\mathbf{f}_{1,7_A}$, the magnitude of $\mathbf{x}_{1_A}$, i.e., $||\mathbf{x}_{1_A}||$, decreases when the line depicting K_A is turned

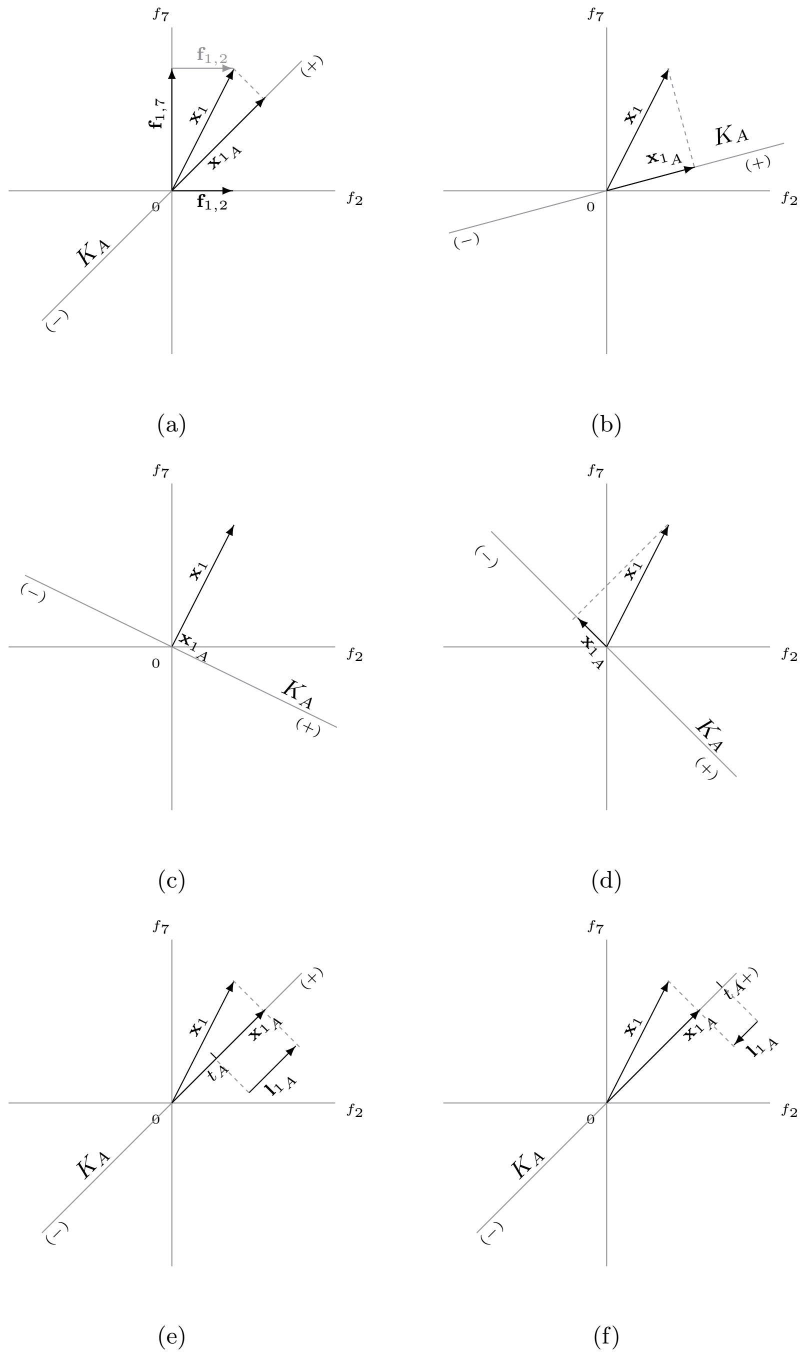


Figure 2.10: Modifying the *resulting* specific influence of *cookie* 1's features.

clockwise as shown in Figure 2.10b. Furthermore, after this line is turned clockwise as depicted in Figure 2.10c, the magnitude of the resulting specific influence of these features disappears, i.e., $||\mathbf{x}_{1_A}|| = 0$, which means that in this case the features do not have any specific influence on the appraisal of *cookie* 1 satisfying $\mathcal{C}$. Figure 2.10d shows a situation in which the position of the line yields a resulting specific influence that disfavors the fulfillment of $\mathcal{C}$, i.e., a situation in which the direction of $\mathbf{x}_{1_A}$ is opposite to the direction of $\hat{\mathbf{u}}_A$.

In Figures 2.10e and 2.10f, it is shown how a variation of the threshold point t_A can also have an effect on the level to which *cookie* 1 satisfies $\mathcal{C}$: while in the situation depicted in Figure 2.10e *cookie* 1 can be labeled as a 'made by Grandma' cookie because the resulting specific influence overcomes t_A, in the situation shown in Figure 2.10f the cookie cannot be labeled as such since the resulting specific influence does not overcome t_A.

The above scenarios suggest that, during the Pia's learning process, the components of K_A, i.e., $\hat{\mathbf{u}}_A$ and t_A, can be varied in order to increase, decrease or even change the direction and the level of the resulting specific influence of the features on the appraisal of $\mathcal{C}$ on a given cookie.

Following the above suggestion, one can try to adjust the components of K_A in such a way that the resulting specific influence of the features of each cookie is in agreement with the label assigned to it in the training data set. For example, consider that, during the Pia's learning process *cookie* 1 is first studied, followed by *cookie* 3. Figure 2.11a shows a possible representation of the knowledge acquired by Pia after studying the first cookie. Notice in this figure that the vector $\mathbf{l}_{1_A}$, which represents the level to which *cookie* 1 satisfies the criterion $\mathcal{C}$, is in agreement with the label assigned to this cookie in the training data set. Hence, one can say that the line K_A represented in Figure 2.11a is in agreement with the assigned label since in the training data set *cookie* 1 has been labeled as a '*made by Grandma'* cookie. However, after studying *cookie* 3 and including its representation in Figure 2.11b, one can notice that the vector $\mathbf{l}_{3_A}$ denotes the fulfillment of C on *cookie* 3, which is in disagreement with the '*not made by Grandma'* label assigned to this cookie in the training data set. Thus, one can say that, in this case, the representation of K_A does not reflect what Pia understands as a Grandma's cookie. To restore the agreement between each cookie and its assigned label, one can adjust the components of K_A as shown in Figure 2.11c. In this case, both vectors $\mathbf{l}_{1_A}$ and $\mathbf{l}_{3_A}$ are in agreement with the labels assigned to them and, thus, one can say that the depiction of K_A in this figure is suitable for representing the Pia's knowledge about what a Grandma's cookie is.

As can be noticed, the idea behind the third step is to find potential suitable directional vectors and threshold points for a given concept, where suitable means that the resulting specific influence of the features of each $x_i \in X_0$ must correspond to label given for x_i in the training data set; thus, e.g., if x_i has been labeled as an object that fulfills a criterion $\mathcal{C}$, the resulting specific influence of its features should be in favor of the fulfillment of $\mathcal{C}$ in such a way that the threshold t_A is exceeded.

After the potential suitable directional vectors and threshold points have

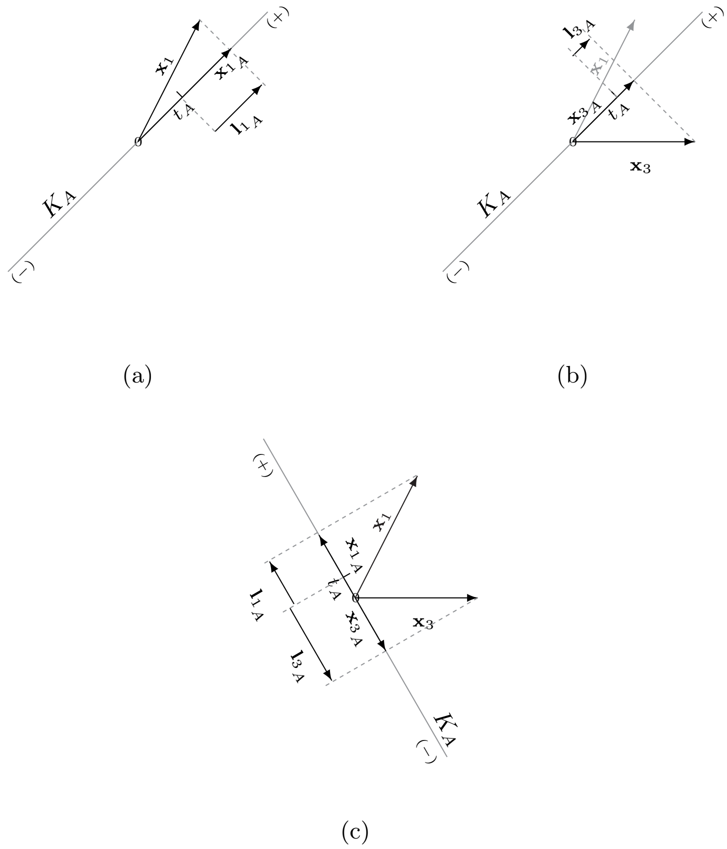


Figure 2.11: Finding suitable directional vectors and threshold points.

been found, one can choose a couple that maximizes both the aggregate of the specific influences of the features in favor of $\mathcal{C}$ and the aggregate of the specific influences of the features in opposition to $\mathcal{C}$. The chosen couple is returned in the last step of the learning process. A method that computes such an optimal couple will be explained in the next section.

2.2.4 Computing the couple $\langle \hat{\mathbf{u}}_A, t_A \rangle$ that represents a particular understanding of concept A

To perform the computation of an optimal couple $\langle \hat{\mathbf{u}}_A, t_A \rangle$, one can use a *support vector machine*, SVM for short [3, 4], which has been successfully used for handling pattern recognition problems in statistical learning theory (see, e.g., the applications for optical character recognition [5], face detection [6], and text categorization [7]).

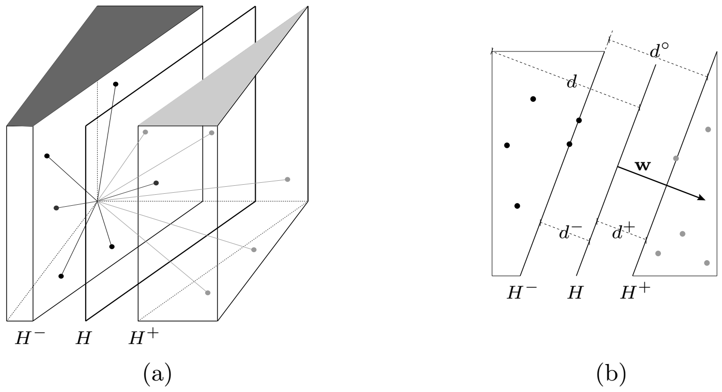


Figure 2.12: Idea behind an SVM: a separating hyperplane H can be used to categorize objects.

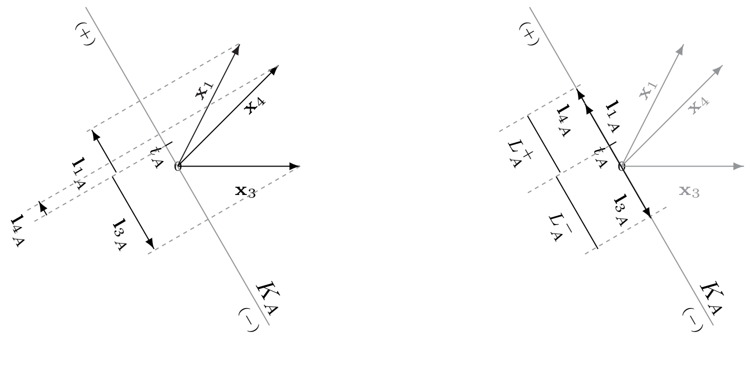
An SVM is a classifier based on the idea that a *separating hyperplane*, i.e., a surface with $m-1$ dimensions that separates a m-dimensional space into two parts, can be used to categorize objects. This idea is depicted in Figure 2.12: while Figure 2.12a shows a 2-dimensional hyperplane H that separates a 3-dimensional space into two parts, one containing black circles and the other including gray circles, Figure 2.12b illustrates such parts in a 2-dimensional space separated by a 1-dimensional hyperplane H. In this regard, finding an optimal hyperplane that separates objects belonging to a given category from others is a problem that an SVM tries to solve.

In what follows, the problem about finding an optimal couple $\langle \hat{\mathbf{u}}_A, t_A \rangle$ and the problem about finding an optimal hyperplane with an SVM are stated. Then, a relation between these problems is explained. Finally, it is shown how an SVM obtains the results.

Let $\mathcal{C}$ be an evaluation criterion related to a concept A. Let $X_0 = \{x_1, \cdots, x_n\}$ and $Y_0 = \{y_1, \cdots, y_n\}$ be constituents of a training data set in which each $x_i \in X_0$ is an object that has been assigned a label $y_i \in Y_0$ according to the following conditions: when x_i satisfies $\mathcal{C}$, i.e., x_i is a *positive example*, y_i is 1; and when x_i dissatisfies $\mathcal{C}$, i.e., x_i is a *negative example*, y_i is -1. Assume that $\mathbf{x}_i$ and $\mathbf{x}_{iA}$ denote respectively the overall and the specific resulting influence of the features of an object x_i in X_0 according to the feature-influence representational model.

With these considerations, the problem of computing an optimal couple $\langle \hat{\mathbf{u}}_A, t_A \rangle$ that characterizes a particular understanding of concept A, say K_A, can be formulated as follows:

> Let $\mathbf{l}_{iA}^+$ and $\mathbf{l}_{iA}^-$ denote vectors that represent the level to which an object x_i in X_0 *satisfies* and *dissatisfies* $\mathcal{C}$ respectively – e.g., in Figure 2.13a, let $\mathbf{l}_{1A}^+$, $\mathbf{l}_{4A}^+$ and $\mathbf{l}_{3A}^-$ denote $\mathbf{l}_{1A}$, $\mathbf{l}_{4A}$ and $\mathbf{l}_{3A}$ respectively. Assume $L_A^+ = \sum_{x_i \in X_0} ||\mathbf{l}_{iA}^+||$ and $L_A^- = \sum_{x_i \in X_0} ||\mathbf{l}_{iA}^-||$ – e.g. in Figure 2.13b, assume both $L_A^+ = ||\mathbf{l}_{1A}^+|| + ||\mathbf{l}_{4A}^+||$ and $L_A^- = ||\mathbf{l}_{3A}^-||$. Find $\hat{\mathbf{u}}_A$ and t_A such that both L_A^+ and L_A^- are maximized.



(a) (b)

Figure 2.13: Finding K_A that maximizes both L_A^+ and L_A^-.

The problem about finding an optimal hyperplane with an SVM is formulated as follows [8]:

> Suppose that a hyperplane H separates vectors $\mathbf{x}_i$ corresponding to positive examples and negative ones – e.g., in Figure 2.12, H separates positive examples represented as gray circles from negative examples represented as black circles. Let H^+ be a hyperplane that is parallel to H and contains the closest positive example(s) to it, and let H^- be another hyperplane that is also parallel to H and contains the closest negative example(s) to it. Let d^+ be the distance between H^+ and H and let d^- be the distance between H^- and H. Finally, let $d^\circ = d^+ + d^-$ be the distance between H^+ and H^-. Find H such that d° is the largest.

The hyperplane H can be defined by a perpendicular (or normal) vector $\mathbf{w}$ and a term b. Considering that in a vector space a point can be denoted by a vector whose tail is the origin and whose head is the point itself, all the points $\mathbf{x}$ on H satisfy $\mathbf{w} \cdot \mathbf{x} + b = 0$ such that $d = |b| / \|\mathbf{w}\|$ is the perpendicular distance from H to the origin, $|b|$ is the absolute value of b, and $\|\mathbf{w}\|$ is the magnitude of $\mathbf{w}$. This case is depicted in Figure 2.12, in which vectors corresponding to objects that satisfy (gray-circle heads) or dissatisfy (black-circle heads) an evaluation criterion are presented in a 3-dimensional space and, also, in a 2-dimensional space. Notice that (i) the hyperplane H separates the positive from the negative examples; (ii) the hyperplane H^+ is parallel to H and contains the closest positive example to it; (iii) the hyperplane H^- is also parallel to H and contains the closest negative example to it; and (iv) the distance $d^\circ = d^+ + d^-$ between H^+ and H^- is the largest.

A relation between the required solutions to the aforementioned problems is shown in Figure 2.14. Notice in this figure that (i) the vector $\mathbf{w}$ points to the side that contains the positive examples; (ii) the normal vector to H, i.e., $\mathbf{w}$, and the *directional vector* $\hat{\mathbf{u}}_A$ are parallel to each other and point to the same

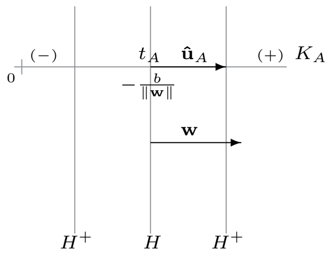


Figure 2.14: An optimal couple $\langle \hat{\mathbf{u}}_A, t_A \rangle$ in relation to an optimal separating hyperplane H.

side; and (iii) intersect term b in a hyperplane $H : \mathbf{w} \cdot \mathbf{x} + b = 0$ corresponds to the *threshold point* t_A on the line that represents A. Hence, the equations

$$\hat{\mathbf{u}}_A = \frac{\mathbf{w}}{\|\mathbf{w}\|} \tag{2.2}$$

and

$$t_A = -\frac{b}{\|\mathbf{w}\|} \tag{2.3}$$

hold. Notice also that turning the hyperplane H will affect the directional vector $\hat{\mathbf{u}}_A$ and, thus, it will affect the resulting specific influence of the features on the learning concept as shown above.

From the previous relation, the specific influence of x_i corresponds to

$$\mathbf{x}_{iA} = \frac{(\mathbf{x}_i \cdot \mathbf{w})}{\|\mathbf{w}\|} \hat{\mathbf{u}}_A, \tag{2.4}$$

and, from (2.1), the vector that denotes the level to which x_i satisfies (or dissatisfies) $\mathcal{C}$ corresponds to

$$\mathbf{l}_{iA} = \frac{(\mathbf{x}_i \cdot \mathbf{w}) + b}{\|\mathbf{w}\|} \hat{\mathbf{u}}_A. \tag{2.5}$$

Having defined the relation between the required solutions, we can now explain how an optimal separating hyperplane is obtained by an SVM – the interested reader is referred to [8] for a detailed tutorial about SVMs.

Consider that all the training vectors $\mathbf{x_i}$, i.e., the vectors corresponding to the objects $x_i \in X_0$, satisfy the following constraints:

- if x_i is a positive example (i.e., $y_i = 1$), then

$$\mathbf{w} \cdot \mathbf{x_i} + b \geq 1; \tag{2.6}$$

- if x_i is a negative example (i.e., $y_i = -1$), then

$$\mathbf{w} \cdot \mathbf{x_i} + b \leq -1. \tag{2.7}$$

Now consider that the perpendicular distance from H^+ to the origin is $|1-b|/\|\mathbf{w}\|$. Consider also that the heads of some vectors $\mathbf{x_i}$ are on H^+, i.e., they satisfy $\mathbf{w}\cdot\mathbf{x}_i+b=1$. Similarly, consider that the perpendicular distance from H^- to the origin is $|-1-b|/\|\mathbf{w}\|$ and that some of the heads of the vectors $\mathbf{x_i}$ are on H^-. Hence, the distance between H and H^+ is $d^+=1/\|\mathbf{w}\|$ and the distance between H and H^- is also $d^-=1/\|\mathbf{w}\|$. This means that the distance d° between H^+ and H^- is $d^\circ=d^++d^-=2/\|\mathbf{w}\|$. Therefore, we can find the hyperplanes H^+ and H^- that maximize d° by maximizing $2/\|\mathbf{w}\|$ subject to the constraints (2.6) and (2.7). The vectors $\mathbf{x_i}$ whose heads are on any of the hyperplanes H^+ and H^- that maximize m are called *support vectors*.

For simplicity, instead of maximizing $2/\|\mathbf{w}\|$, minimizing $\frac{1}{2}\|\mathbf{w}\|^2$ is preferred [8]. Thus, the problem of finding an optimal separating hyperplane can be reformulated as the following minimization problem:

> Find $\mathbf{w}$ and b such that $\frac{1}{2}\|\mathbf{w}\|^2$ is minimized and the constraints (2.6) and (2.7) hold for each vector $\mathbf{x_i}$.

This is a problem in which a quadratic function is optimized subject to linear constraints. The solution involves first switching to the following Lagrangian formulation of the problem [8]:

> Let $\lambda_1,\cdots,\lambda_n$ be positive Lagrange multipliers, where a multiplier λ_i is associated to each $y_i(\mathbf{w}\cdot\mathbf{x_i}+b)-1\geq 0$, which is a "compact" version of the constraints (2.6) and (2.7); and let Λ be a Lagrangian defined by
>
> $$\Lambda=\frac{1}{2}\|\mathbf{w}\|^2-\sum_{i=1}^{n}\lambda_i\left(y_i(\mathbf{w}\cdot\mathbf{x}_i+b)-1\right). \tag{2.8}$$
>
> Find $\mathbf{w}$, b and all the λ_i such that Λ is minimized.

This problem is then reformulated to the following equivalent dual problem [8]:

> Find $\lambda_1,\cdots,\lambda_n$ such that the gradient of Λ with respect to $\mathbf{w}$ and b yields zero, and Λ is maximized.

The condition for the gradient of Λ results in

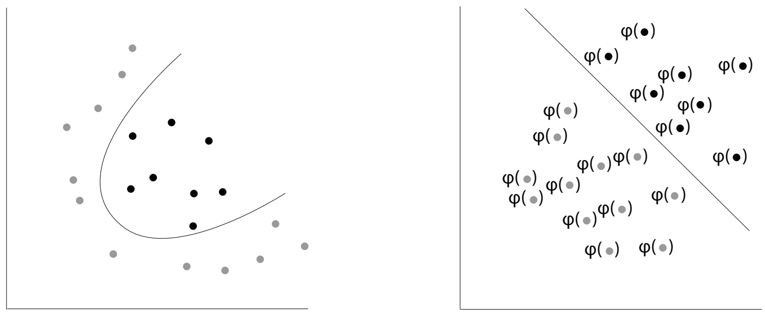
$$\mathbf{w}=\sum_{i=1}^{n}\lambda_i y_i\mathbf{x}_i \tag{2.9}$$

and

$$\sum_{i=1}^{n}\lambda_i y_i=0, \tag{2.10}$$

which are replaced in (2.8) to obtain

$$\Lambda=\sum_{i=1}^{n}\lambda_i-\frac{1}{2}\sum_{i=1,k=1}^{n}\lambda_i\lambda_k y_i y_k(\mathbf{x}_i\cdot\mathbf{x}_k). \tag{2.11}$$



(a) (b)

Figure 2.15: Dealing with non linearly separable objects.

Notice that in this formulation Λ is maximized with respect to the multipliers λ_i subject to (2.10) and $\lambda_i \geq 0$. The solution here is given by both (2.9) and

$$b = y_i - \mathbf{w} \cdot \mathbf{x}_i, \tag{2.12}$$

for any $\mathbf{x}_i$ such that $\lambda_i > 0$. Notice also that there is a multiplier λ_i for each $\mathbf{x}_i$. In the solution, the vectors $\mathbf{x}_i$ having multipliers $\lambda_i > 0$ are the *support vectors*, which can be deemed to be crucial elements of the training data set because these vector can change the location of the separating hyperplane if removed – in [9], a software tool called *SVMLight* has been provided to perform the computation of both (2.9) and (2.12).

At this point, the careful reader may realize that the above solution is based on a situation in which objects belonging to a given category can be *linearly* separated from others not belonging to the category. Thus, this person might ask: *what happens in a situation like the one depicted in Figure 2.15a where the objects are not linearly separable?*

In such a situation, one can map those objects to another space in which they can be separated by a linear hyperplane as shown in Figure 2.15b. In other words, one can map a vector $\mathbf{x}_i$ in the feature space $\mathcal{M}$ to a higher dimensional space, say $\mathcal{N}$, through a mapping ϕ such that $\phi : \mathcal{M} \mapsto \mathcal{N}$. Hence (2.11) becomes

$$\Lambda = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1,k=1}^{n} \lambda_i \lambda_k y_i y_k (\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_k)). \tag{2.13}$$

Since computing the inner product between $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_k)$ in $\mathcal{N}$ can add more complexity, it is preferred to use a *kernel function*, say K, such that $K(\mathbf{x}_i, \mathbf{x}_k) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_k)$ [3, 4] – i.e., K computes the inner product (or an interpretation of similarity) between $\mathbf{x}_i$ and $\mathbf{x}_k$ in $\mathcal{N}$. Hence, (2.13) can be rewritten as

$$\Lambda = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1,k=1}^{n} \lambda_i \lambda_k y_i y_k (K(\mathbf{x}_i, \mathbf{x}_k)), \tag{2.14}$$

which does not depend of the aforementioned mapping. Some examples of kernel functions are the following:

- *Linear kernel,* defined by

$$K(\mathbf{x}_i, \mathbf{x}_k) = (\mathbf{x}_i \cdot \mathbf{x}_k)\,; \tag{2.15}$$

- *Polynomial kernel* of degree p, defined by

$$K(\mathbf{x}_i, \mathbf{x}_k) = (\mathbf{x}_i \cdot \mathbf{x}_k + 1)^p\,; \tag{2.16}$$

- *Radial Basis Function (RBF) kernel* with a parameter $\gamma > 0$, defined by

$$K(\mathbf{x}_i, \mathbf{x}_k) = \exp\left(-\gamma ||\mathbf{x}_i - \mathbf{x}_k||^2\right); \text{ and} \tag{2.17}$$

- *Sigmoid kernel* with parameters $\kappa > 0$ and $\delta \in \mathbb{R}$, defined by

$$K(\mathbf{x}_i, \mathbf{x}_k) = \tanh(\kappa \mathbf{x}_i \cdot \mathbf{x}_k - \delta)\,. \tag{2.18}$$

As can be noticed, an SVM can be used to perform the computation of an optimal couple $\langle \hat{\mathbf{u}}_A, t_A \rangle$ even when the vectors in the training data set are not linearly separable.

As an example, consider the training data set listed in Table 2.2 consisting of the labels and the vectors representing the overall influence of the features detected by Pia in cookies 1, 2, 3, 4 and 5 − see Table 2.1 and recall that, according to the feature-influence representational model, the resulting overall influence corresponds to $\mathbf{x}_i = \beta_{i,1}\hat{\mathbf{f}}_1 + \cdots + \beta_{i,m}\hat{\mathbf{f}}_m$, i.e., the vector sum of all the overall influences of the features in $\mathcal{F}$.

Table 2.2: Pia's learning data set.

Pia's training data set		
cookie (x_i)	vector ($\mathbf{x}_i$)	label (y_i)
$x_1 =$ *cookie* 1	$\mathbf{x}_1 = 1\hat{\mathbf{f}}_2 + 2\hat{\mathbf{f}}_7$	$y_1 = 1$
$x_2 =$ *cookie* 2	$\mathbf{x}_2 = 1\hat{\mathbf{f}}_2 + 2\hat{\mathbf{f}}_3 + 2\hat{\mathbf{f}}_7$	$y_2 = 1$
$x_3 =$ *cookie* 3	$\mathbf{x}_3 = 1\hat{\mathbf{f}}_1 + 2\hat{\mathbf{f}}_3 + 2\hat{\mathbf{f}}_5$	$y_3 = -1$
$x_4 =$ *cookie* 4	$\mathbf{x}_4 = 1\hat{\mathbf{f}}_4 + 2\hat{\mathbf{f}}_6 + 2\hat{\mathbf{f}}_7$	$y_4 = 1$
$x_5 =$ *cookie* 5	$\mathbf{x}_5 = 1\hat{\mathbf{f}}_1 + 2\hat{\mathbf{f}}_6 + 2\hat{\mathbf{f}}_7$	$y_5 = 1$

Using the software tool *SVMLight* with the linear kernel (see (2.15)), one can compute the following multipliers for each $\mathbf{x}_i$ in the Pia's training data set: $\lambda_1 = 0$ ($\mathbf{x}_1$), $\lambda_2 = 0.167$ ($\mathbf{x}_2$), $\lambda_3 = 0.208$ ($\mathbf{x}_3$), $\lambda_4 = 0$ ($\mathbf{x}_4$) and $\lambda_5 = 0.042$ ($\mathbf{x}_5$). Since λ_2, λ_3 and λ_5 are greater than 0, one can say that the vectors corresponding to cookies 2, 3 and 5 constitute the support vectors in this case.

One can also obtain the following values of the vector $\mathbf{w}$ and the term b that define the separating hyperplane H: $\mathbf{w} = -0.167\hat{\mathbf{f}}_1 + 0.167\hat{\mathbf{f}}_2 - 0.083\hat{\mathbf{f}}_3 - 0.417\hat{\mathbf{f}}_5 + 0.083\hat{\mathbf{f}}_6 + 0.417\hat{\mathbf{f}}_7$ and $b = 0.167$.

Using the values of $\mathbf{w}$ and b in (2.2) and (2.3), one can compute the couple $\langle \hat{\mathbf{u}}_{A@Pia}, t_{A@Pia} \rangle$, which represents what Pia learned about the *Grandma's cookie* concept. The results are

$$\begin{aligned} \hat{\mathbf{u}}_{A@Pia} &= -0.258\hat{\mathbf{f}}_1 + 0.258\hat{\mathbf{f}}_2 - 0.129\hat{\mathbf{f}}_3 \\ &\qquad - 0.645\hat{\mathbf{f}}_5 + 0.129\hat{\mathbf{f}}_6 + 0.645\hat{\mathbf{f}}_7 \text{ and} \\ t_{A@Pia} &= -0.258. \end{aligned}$$

Notice in the components of $\hat{\mathbf{u}}_{A@Pia}$ that, despite of both features favor the fulfillment of the criterion, '*square-shape*' (f_7) is around five times more influential than '*square-hole*' (f_6). Notice also that the magnitude of the influence of '*round-shape*' (f_5) is the same as '*square-shape*' (f_7), but their directions are opposite to each other. Finally, notice that '*curved-icing*' (f_4) has no influence on the fulfillment of the criterion. Since those components could influence on the conditions that arise when an evaluation process is carried out by Pia, (her knowledge) $K_{A@Pia}$ could be deemed to be influential in the context of her (forthcoming) evaluations.

In the next section, it will be explained how such results can be used to mimic an experience-based evaluation process.

2.3 An Experience-Based Evaluation Process

In this section, the question raised is about evaluation: *How can the experience (or knowledge) acquired by a person be reflected in his/her evaluations?* To find an answer to this question, one can use a variant of the intuition given in Section 2.2: *after experiencing with objects that satisfy or dissatisfy an evaluation criterion related to a concept, one obtains a particular knowledge that could be used to appraise the level to which other (new) objects satisfy or dissatisfy the given evaluation criterion.* A process that mimics such a human behavior while evaluating objects in a collection is studied below.

Let $\mathcal{C}$ be a criterion having a form like "be compatible with the way in which A is perceived," where A is a given concept – e.g. let $\mathcal{C}$ be the criterion "be compatible with the way in which *a Grandma's cookie* is perceived." Let $X_0 = \{x_0, \cdots x_n\}$ and $Y_0 = \{y_0, \cdots y_n\}$ be components of a particular training data set used to learn about A as described in the previous section – e.g. let $X_{0@Pia} = \{\textit{cookie } 1, \textit{cookie } 2, \textit{cookie } 3, \textit{cookie } 4, \textit{cookie } 5\}$ and $Y_{0@Pia} = \{1, 1, -1, 1, 1\}$ be the constituents of the training data set used by Pia to learn about what a Grandma's cookie is. Let $\langle \hat{\mathbf{u}}_A, t_A \rangle$ be an optimal couple that represents the knowledge K_A acquired after learning about (concept) A with the aforementioned training data set – e.g., let $\langle \hat{\mathbf{u}}_{A@Pia}, t_{A@Pia} \rangle$ be the optimal couple that represents the knowledge $K_{A@Pia}$ acquired after learning about Grandma's cookies with $X_{0@Pia}$ and $Y_{0@Pia}$. Finally, let X be a collection of objects that are part of an evaluation request (this collection is also known as *test data set*) – e.g., let X be a collection consisting of *cookie* 6, *cookie* 7 and *cookie* 8 (see Figure 2.2).

With these considerations, the level to which an object $x_i \in X$ fulfills $\mathcal{C}$ corresponds to the level to which the *resulting specific influence* of its features on the appraisal of $\mathcal{C}$ (i.e., $\mathbf{x}_{iA}$) exceeds or not (the threshold) t_A according to (the knowledge) K_A. In other words, such an evaluation process consists in the computation of $\mathbf{l}_{iA} = \mathbf{x}_{iA} - t_A\hat{\mathbf{u}}_A$ (see (2.1)) for each $x_i \in X$, where

$$\mathbf{x}_{iA} = (\mathbf{x}_i \cdot \hat{\mathbf{u}}_A)\,\hat{\mathbf{u}}_A, \tag{2.19}$$

and $\mathbf{x}_i$ represents the *resulting overall influence* of (the features of) x_i on the appraisal of $\mathcal{C}$.

For example, to compute the level to which *cookie* 6, *cookie* 7 and *cookie* 8 (see Figure 2.2) satisfy (or dissatisfy) the criterion "be compatible with the way in which *a Grandma's cookie* is perceived" according to Pia's knowledge, i.e., $K_{A@Pia}$, one can consider that the features detected in cookies x_6, x_7 and x_8 are

$$\begin{aligned}
\mathcal{F}_6 &= \{\textit{round-square-shape}, \textit{linear-icing}, \textit{square-hole}\},\\
\mathcal{F}_7 &= \{\textit{round-shape}, \textit{no-icing}, \textit{round-hole}\}, \text{ and}\\
\mathcal{F}_8 &= \{\textit{round-shape}, \textit{no-icing}\}
\end{aligned}$$

respectively. Thus, using the collection of features identified by Pia (see Table 2.1), the vectors that represent cookies x_6, x_7 and x_8 are $\mathbf{x}_{6@Pia} = 1\hat{\mathbf{f}}_1 + 2\hat{\mathbf{f}}_6 + 0\hat{\mathbf{f}}_*$, $\mathbf{x}_{7@Pia} = 1\hat{\mathbf{f}}_2 + 2\hat{\mathbf{f}}_3 + 2\hat{\mathbf{f}}_5$ and $\mathbf{x}_{8@Pia} = 1\hat{\mathbf{f}}_2 + 2\hat{\mathbf{f}}_5$ respectively (see Table 2.3). Notice here that since the feature f_*='*round-square-shape*' was not learned, its neutrality on the appraisal of $\mathcal{C}$ is assumed and, thus, its overall weight is fixed to 0 (its overall influence has been denoted by $0\hat{\mathbf{f}}_*$ for illustration).

Table 2.3: Pia's test data set.

Pia's test data set	
cookie (x_i)	vector ($\mathbf{x}_i$)
x_6 = *cookie* 6	$\mathbf{x}_{6@Pia} = 1\hat{\mathbf{f}}_1 + 2\hat{\mathbf{f}}_6 + 0\hat{\mathbf{f}}_*$
x_7 = *cookie* 7	$\mathbf{x}_{7@Pia} = 1\hat{\mathbf{f}}_2 + 2\hat{\mathbf{f}}_3 + 2\hat{\mathbf{f}}_5$
x_8 = *cookie* 8	$\mathbf{x}_{8@Pia} = 1\hat{\mathbf{f}}_2 + 2\hat{\mathbf{f}}_5$

After representing the cookies as vectors, one can use the evaluation process with, e.g., $\mathbf{x}_{7@Pia}$ to compute the *resulting specific influence* of the features of x_7 as follows:

$$\begin{aligned}
\mathbf{x}_{7A@Pia} &= (\mathbf{x}_{7@Pia} \cdot \hat{\mathbf{u}}_{A@Pia})\hat{\mathbf{u}}_{A@Pia},\\
&= (1 \times (0.258) + 2 \times (-0.129) + 2 \times (-0.645))\,\hat{\mathbf{u}}_{A@Pia}\\
&= -1.29\hat{\mathbf{u}}_{A@Pia}.
\end{aligned}$$

This means that the level to which x_7 satisfies (or dissatisfies) $\mathcal{C}$ is given by

$$\begin{aligned}\mathbf{l}_{7A@Pia} &= -1.29\hat{\mathbf{u}}_{A@Pia} - (-0.258)\hat{\mathbf{u}}_{A@Pia} \\ &= -1.032\hat{\mathbf{u}}_{A@Pia}.\end{aligned}$$

Since $\mathbf{l}_{7A@Pia}$ and $\hat{\mathbf{u}}_{A@Pia}$ have opposite directions, one can say that, according to Pia's knowledge, the cookie x_7 dissatisfies the criterion "be compatible with the way in which *a Grandma's cookie* is perceived" at a level given by $\|\mathbf{l}_{7A@Pia}\| = 1.032$. The levels to which x_6 and x_8 satisfy (or dissatisfy) $\mathcal{C}$ can be computed through the same procedure. The results are shown in Table 2.4.

Table 2.4: Pia's evaluations.

x_i	$\mathbf{l}_{iA@Pia}$	Meaning (according to Pia's understanding)
x_6	$0.258\hat{\mathbf{u}}_{A@Pia}$	x_6 satisfies the criterion at a level 0.258.
x_7	$-1.032\hat{\mathbf{u}}_{A@Pia}$	x_7 dissatisfies the criterion at a level 1.032.
x_8	$-0.775\hat{\mathbf{u}}_{A@Pia}$	x_8 dissatisfies the criterion at a level 0.775.

A visual representation of the above results is shown in Figure 2.16. Notice that the magnitude of the vector representing the level to which *cookie* 6 satisfies the criterion, i.e., $||\mathbf{l}_{6_{A@Pia}}||$, is more or less a quarter of the magnitude of $\mathbf{l}_{7_{A@Pia}}$, which has the largest magnitude among the resulting vectors. Hence, one can use this relation to state: while *cookie* 7 *dissatisfies* the criterion "be compatible with the way in which *a Grandma's cookie* is perceived," *cookie* 6 *slightly satisfies* this criterion. Similarly, one can also state that, in comparison to *cookie* 7, *cookie* 8 *fairly dissatisfies* the criterion.

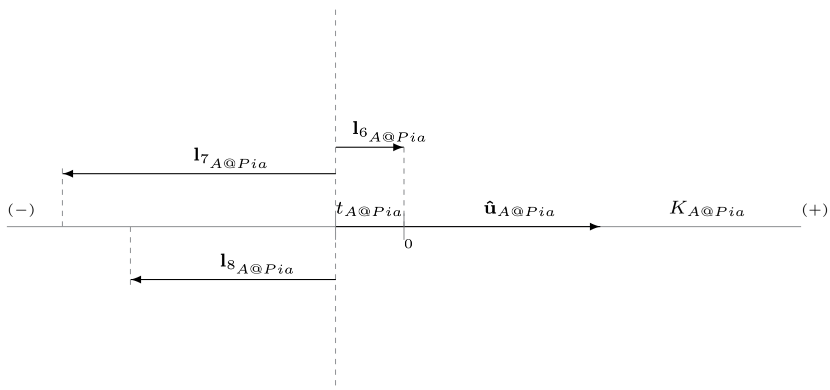


Figure 2.16: A visual representation of Pia's evaluations.

Regarding the aspects that might arise during the evaluation of a cookie, notice in the computation of $\mathbf{x}_{7A@Pia}$ how the knowledge acquired by Pia about the *round shape* (f_5) arises while the evaluation of this cookie satisfying the aforementioned criterion is carried out. In the next section, we present an example that illustrates how other (new) aspects can arise according to the knowledge acquired by other persons about the (same) *Grandma's cookie* concept.

2.4 An Illustrative Example

Although *Pia* and *Rod* can follow the same rules (i.e., the same learning method) to learn which cookies satisfy (or dissatisfy) the criterion "be compatible with the way in which a concept A, say *Grandma's cookies*, is perceived," they can experienced so through different training data sets, say $\{X_{0@Pia}, Y_{0@Pia}\}$ and $\{X_{0@Rod}, Y_{0@Rod}\}$ respectively (see Figure 2.17). Consequently, *Pia* and *Rod* can obtain $K_{A@Pia}$ and $K_{A@Rod}$ in that order as representations of their individual understandings about A. When *Pia* and *Rod* are asked to evaluate the criterion in each element from a new collection, say X, using the same rules (i.e., the same evaluation method) and according to their individual understandings, they can provide the collections $Y_{@Pia}$ and $Y_{@Rod}$ containing respectively their individual XBEs.

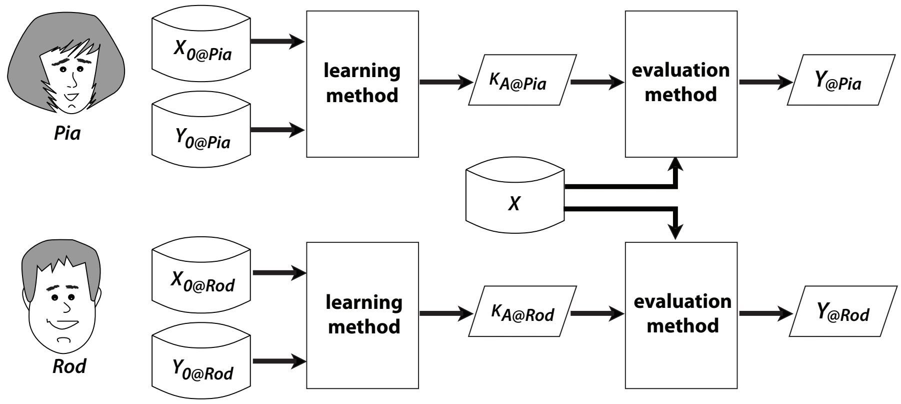


Figure 2.17: Learning and evaluation processes followed by Pia and Rod.

With those considerations, among others, a difference between $Y_{@Pia}$ and $Y_{@Rod}$ is mainly determined by a difference between $K_{A@Pia}$ and $K_{A@Rod}$. To illustrate so, in what follows we make use of the learning method described in Section 2.2 to simulate, as was done with Pia, a learning process followed by Rod. After that, we use the evaluation method described in the previous section, to mimic the evaluation process followed by Rod when he is asked to evaluate to which degree each cookie in $X = \{x_6, x_7, x_8\}$ is a Grandma's cookie (i.e., Pia and Rod received the same evaluation request).

Rod's learning process. To learn about a Grandma's cookie, Rod can study, as Pia did, the features of the cookies that satisfy or dissatisfy the criterion "be compatible with the way in which a *Grandma's cookie* is perceived." To do so, he can use a training data set including the cookies given by his mom and the labels indicating whether or not a cookie is a Grandma's cookie (see Example 2.1). These cookies and labels can be represented as $X_{0@Rod} = \{x_1, x_2, x_3, x_4, x_7\}$ and $Y_{0@Rod} = \{y_1, y_2, y_3, y_4, y_7\}$ respectively – recall that when a cookie, say x_i, is made by Grandma, a label, say y_i, with a value of 1 is assigned to this cookie, and when x_i is not made by Grandma, a label $y_i = -1$ is assigned. The features detected in cookies x_1, x_2, x_3, x_4, x_7, i.e., the cookies in $X_{0@Rod}$, are

$$\begin{aligned}
\mathcal{F}_1 &= \{\textit{square-shape}, \textit{no-icing}\}, \\
\mathcal{F}_2 &= \{\textit{square-shape}, \textit{no-icing}, \textit{round-hole}\}, \\
\mathcal{F}_3 &= \{\textit{round-shape}, \textit{linear-icing}, \textit{round-hole}\}, \\
\mathcal{F}_4 &= \{\textit{square-shape}, \textit{curved-icing}, \textit{square-hole}\}, \text{ and} \\
\mathcal{F}_7 &= \{\textit{round-shape}, \textit{no-icing}, \textit{round-hole}\}
\end{aligned}$$

respectively. The collection $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \cup \mathcal{F}_4 \cup \mathcal{F}_7$ is listed in the first column of Table 2.5.

Table 2.5: Collection of features identified by Rod.

Rod		
feature	f_j	overall weight
linear-icing	f_1	3
no-icing	f_2	3
round-hole	f_3	1
curved-icing	f_4	3
round-shape	f_5	1
square-hole	f_6	1
square-shape	f_7	1

To reflect what Rod's mom told him about the preferences of Grandma, i.e., making cookies without icing, he can consider that the features related to the icing are three times more important than the others. If so, he can assign the overall influence of each feature as shown in the third column of Table 2.5 – as will be shown in Section 4.3.1, the overall influence of a feature can also be automatically assigned according to a heuristic process.

It is worth mentioning that Rod can consider the features related to icing twice more important than the others, like Pia did. Even more, he could consider that such features are ten times more important, i.e., he could consider a 10 to 1 (or even a 100 to 1) ratio between the most and the least important features. However, choosing such a high ratio can make the influence of the least important features practically *disappear* during the learning process. In other words, choosing such a high ratio is like having cookies in which only the most important features are present. To visualize this, several mental pictures (or abstractions) of *cookie* 3 that Rod might have according to the overall influence assigned to the icing are depicted in Figure 2.18: in (a), the importance of the icing is deemed to be similar to the other features in *cookie* 3 (1 to 1 ratio); in (b), the icing is deemed to be three times more important than the other features (3 to 1 ratio); and in (c), the icing is considered to be ten times more important than the others (10 to 1 ratio). As could be noticed, the overall influence assigned to the icing will have an significant effect on the abstraction of a cookie like *cookie* 3 during a learning process.

(a) 1 : 1. (b) 3 : 1. (c) 10 : 1.

Figure 2.18: Setting the overall influence of a feature.

After assigning the overall influence to each feature, one can represent the training data set used by Rod by means of the feature-influence representational model as shown in Table 2.6.

Table 2.6: Rod's learning data set.

Rod's training data set		
cookie (x_i)	vector ($\mathbf{x}_i$)	label (y_i)
$x_1 = \textit{cookie}\ 1$	$\mathbf{x}_1 = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_7$	$y_1 = 1$
$x_2 = \textit{cookie}\ 2$	$\mathbf{x}_2 = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_3 + 1\hat{\mathbf{f}}_7$	$y_2 = 1$
$x_3 = \textit{cookie}\ 3$	$\mathbf{x}_3 = 3\hat{\mathbf{f}}_1 + 1\hat{\mathbf{f}}_3 + 1\hat{\mathbf{f}}_5$	$y_3 = -1$
$x_4 = \textit{cookie}\ 4$	$\mathbf{x}_4 = 3\hat{\mathbf{f}}_4 + 1\hat{\mathbf{f}}_6 + 1\hat{\mathbf{f}}_7$	$y_4 = -1$
$x_7 = \textit{cookie}\ 7$	$\mathbf{x}_7 = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_3 + 1\hat{\mathbf{f}}_5$	$y_7 = 1$

As was done during the learning process followed by Pia, one can follow the procedure described in Section 2.2.4 with the training data set shown in Table 2.6 to compute the couple $\langle \hat{\mathbf{u}}_{A@Rod}, t_{A@Rod} \rangle$, which represents what Rod learned about the *Grandma's cookie* concept, i.e., $K_{A@Rod}$. In this case, the results that characterize $K_{A@Rod}$ are the following:

$$\hat{\mathbf{u}}_{A@Rod} = -0.426\hat{\mathbf{f}}_1 + 0.809\hat{\mathbf{f}}_2 - 0.383\hat{\mathbf{f}}_4 - 0.128\hat{\mathbf{f}}_6 \text{ and}$$
$$t_{A@Rod} = 0.575.$$

Notice in the components of $\hat{\mathbf{u}}_{A@Rod}$ that, while the features '*linear-icing*' (f_1) and '*curved-icing*' (f_4) disfavor the fulfillment of the criterion, the '*no-icing*' aspect (f_2) favors so. Notice also that, in contrast to the representation of

Pia's knowledge, the feature '*square-hole*' (f_6) disfavors the fulfillment of the criterion.

Rod's evaluation process. Regarding the evaluation process, to compute the level to which each cookie in the test data set, i.e., *cookie* 6, *cookie* 7 and *cookie* 8 (see Figure 2.2), satisfies (or dissatisfies) the criterion "be compatible with the way in which *a Grandma's cookie* is perceived" according to Rod's knowledge, i.e., $K_{A@Rod}$, one can consider that the features detected in cookies x_6, x_7 and x_8 are the same detected by Pia, i.e.,

$$\begin{aligned}
\mathcal{F}_6 &= \{\textit{round-square-shape}, \textit{linear-icing}, \textit{square-hole}\}, \\
\mathcal{F}_7 &= \{\textit{round-shape}, \textit{no-icing}, \textit{round-hole}\}, \text{ and} \\
\mathcal{F}_8 &= \{\textit{round-shape}, \textit{no-icing}\}
\end{aligned}$$

respectively. However, to represent cookies x_6, x_7 and x_8 as vectors in this case, one must use the collection of features identified by Rod (see Table 2.5). Hence, one can represent these cookies as $\mathbf{x}_{6@Rod} = 3\hat{\mathbf{f}}_1 + 1\hat{\mathbf{f}}_6 + 0\hat{\mathbf{f}}_*$, $\mathbf{x}_{7@Rod} = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_3 + 1\hat{\mathbf{f}}_5$ and $\mathbf{x}_{8@Rod} = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_5$ respectively (see Table 2.7). Notice here that, in a similar way to the case of Pia, since the feature f_*='*round-square-shape*' was not learned, its neutrality on the appraisal of $\mathcal{C}$ is assumed.

Table 2.7: Rod's test data set.

Rod's test data set	
cookie (x_i)	vector ($\mathbf{x}_i$)
$x_6 =$ *cookie* 6	$\mathbf{x}_{6@Rod} = 3\hat{\mathbf{f}}_1 + 1\hat{\mathbf{f}}_6 + 0\hat{\mathbf{f}}_*$
$x_7 =$ *cookie* 7	$\mathbf{x}_{7@Rod} = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_3 + 1\hat{\mathbf{f}}_5$
$x_8 =$ *cookie* 8	$\mathbf{x}_{8@Rod} = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_5$

After representing the cookies as vectors according to Rod's perspective, one can use the evaluation process to compute the resulting specific influence of each cookie. For instance, one can compute resulting specific influence of *cookie* 7 as follows:

$$\begin{aligned}
\mathbf{x}_{7A@Rod} &= (\mathbf{x}_{7@Rod} \cdot \hat{\mathbf{u}}_{A@Rod})\hat{\mathbf{u}}_{A@Rod}, \\
&= (3 \times (0.809) + 1 \times (0) + 1 \times (0))\,\hat{\mathbf{u}}_{A@Rod} \\
&= 2.427\hat{\mathbf{u}}_{A@Rod}.
\end{aligned}$$

This means that the level to which x_7 satisfies (or dissatisfies) $\mathcal{C}$ is given by

$$\begin{aligned}
\mathbf{l}_{7A@Rod} &= 2.427\hat{\mathbf{u}}_{A@Rod} - (0.575)\hat{\mathbf{u}}_{A@Rod} \\
&= 1.852\hat{\mathbf{u}}_{A@Rod}.
\end{aligned}$$

Since $\mathbf{l}_{7A@Rod}$ and $\hat{\mathbf{u}}_{A@Rod}$ have the same direction, one can say that, according to Rod's knowledge, the cookie denoted by x_7 satisfies the criterion "be compatible with the way in which *a Grandma's cookie* is perceived" at a level given by $\|\mathbf{l}_{7A@Rod}\| = 1.852$. Notice in this computation that the features *round-hole* (f_3) and *round-shape* (f_5) have no influence on the evaluation. The resulting specific influences of *cookie* 6 and *cookie* 8 are shown in Table 2.8.

Table 2.8: Rod's evaluations.

x_i	$\mathbf{l}_{iA@Rod}$	Meaning (according to Rod's understanding)
x_6	$-1.981\hat{\mathbf{u}}_{A@Rod}$	x_6 dissatisfies the criterion at a level 1.981.
x_7	$1.852\hat{\mathbf{u}}_{A@Rod}$	x_7 satisfies the criterion at a level 1.852.
x_8	$1.852\hat{\mathbf{u}}_{A@Rod}$	x_8 satisfies the criterion at a level 1.852.

In the visual representation of these XBEs shown in Figure 2.19, one can notice that the magnitude of the levels to which each cookie in the test set satisfies (or dissatisfies) the criterion are practically the same. Therefore, one can state that, while *cookie* 6 *dissatisfies* the criterion"be compatible with the way in which a Grandma's cookie is perceived," both *cookie* 7 and *cookie* 8 *satisfy* so.

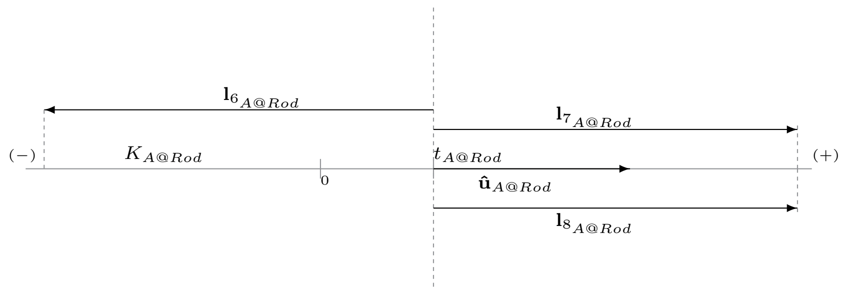


Figure 2.19: A visual representation of Rod's evaluations.

At this point, one can realize that the XBEs of cookies 6, 7 and 8 based on Rod's knowledge, i.e., $K_{A@Rod}$, differ from the XBEs of these cookies based on Pia's knowledge, i.e., $K_{A@Pia}$. Since both the learning process and the evaluation process are the same, any difference Pia's and Rod's XBEs is mainly determined by any difference between $K_{A@Pia}$ and $K_{A@Rod}$. In other words, any difference in the XBEs is mainly determined by a difference in individual understandings that Pia and Rod have about the Grandma's cookie concept. To visualize this, $K_{A@Pia}$ and $K_{A@Rod}$ are depicted in Figure 2.20. Observe that the lines representing to $K_{A@Pia}$ and $K_{A@Rod}$ are not aligned to each other and, also, the threshold points are located in different positions.

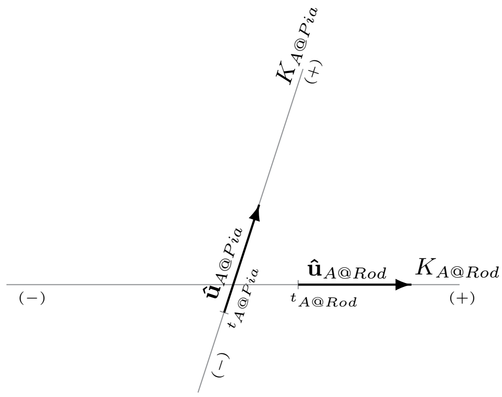


Figure 2.20: A graphical view of the difference in understandings about the Grandma's cookie concept between Pia and Rod.

Recalling from Section 1.4.1 that a group of respondents (or evaluators) can be deemed to be *homogeneous* when its members have a very similar (or the same) understanding of the concept under study, one can say that Pia and Rod constitute a *heterogeneous group* since their understandings about Grandma's cookies are dissimilar to each other. Here, someone may ask a natural question: *how aligned (or similar) such understandings are?* To get insights on this regard, one can consider the following situation in which opposite examples are used in a training data set:

> Rod challenged his friend Sam to learn about Grandma's cookies by using opposite examples in his training data set – by opposite example is meant that, e.g., if a cookie is deemed to be a Grandma's cookie in the training data set used by Rod, the cookie will not considered as such in the training data set used by Sam. In addition, Sam has been asked to use the same features detected by Rod. *What can Sam learn about a Grandma's cookie from the modified training data set?*

In this case, since the collection of features is the same collection used by Rod (see Table 2.5), the training data set for Sam's learning process is the same as Rod but with opposite labels, as shown in Table 2.9.

After mimicking the learning process followed by Sam, we obtain $K_{A@Sam}$, which is characterized by

$$\hat{\mathbf{u}}_{A@Sam} = 0.426\hat{\mathbf{f}}_1 - 0.809\hat{\mathbf{f}}_2 + 0.383\hat{\mathbf{f}}_4 + 0.128\hat{\mathbf{f}}_6 \text{ and}$$
$$t_{A@Sam} = -0.575.$$

As could be expected, after using a training data set with opposite examples $\hat{\mathbf{u}}_{A@Sam} = -\hat{\mathbf{u}}_{A@Rod}$ holds, which reflects a totally opposite understanding of the Grandma's cookie in comparison to Rod's. This opposite understanding is then reflected when we use $K_{A@Sam}$ to evaluate to which degree each cookie in

Table 2.9: Sam's learning data set.

Sam's training data set		
cookie (x_i)	vector ($\mathbf{x}_i$)	label (y_i)
$x_1 = cookie\ 1$	$\mathbf{x}_1 = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_7$	$y_1 = -1$
$x_2 = cookie\ 2$	$\mathbf{x}_2 = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_3 + 1\hat{\mathbf{f}}_7$	$y_2 = -1$
$x_3 = cookie\ 3$	$\mathbf{x}_3 = 3\hat{\mathbf{f}}_1 + 1\hat{\mathbf{f}}_3 + 1\hat{\mathbf{f}}_5$	$y_3 = 1$
$x_4 = cookie\ 4$	$\mathbf{x}_4 = 3\hat{\mathbf{f}}_4 + 1\hat{\mathbf{f}}_6 + 1\hat{\mathbf{f}}_7$	$y_4 = 1$
$x_7 = cookie\ 7$	$\mathbf{x}_7 = 3\hat{\mathbf{f}}_2 + 1\hat{\mathbf{f}}_3 + 1\hat{\mathbf{f}}_5$	$y_7 = -1$

$X = \{x_6, x_7, x_8\}$ is a Grandma's cookie. The resulting evaluations are shown in Table 2.10. Notice how the opposition in understandings is also evident in these evaluations: while in Rod's evaluations the level to which, e.g., cookie 8 *satisfies* the Grandma's cookie criterion is 1.852 (mostly) because of the *no-icing* feature, in Sam's evaluations the level to which this cookie *dissatisfies* the criterion is 1.852 because of the same feature.

Table 2.10: Sam's evaluations.

x_i	$\mathbf{l}_{iA@Sam}$	Meaning (according to Sam's understanding)
x_6	$1.981\hat{\mathbf{u}}_{A@Sam}$	x_6 satisfies the criterion at a level 1.981.
x_7	$-1.852\hat{\mathbf{u}}_{A@Sam}$	x_7 dissatisfies the criterion at a level 1.852.
x_8	$-1.852\hat{\mathbf{u}}_{A@Sam}$	x_8 dissatisfies the criterion at a level 1.852.

Since Sam's understanding about Grandma's cookies is completely opposite to Rod's, one can use it to get an estimate of how aligned Rod's and Pia's understandings are. For instance, in the visual representation of their understandings shown in Figure 2.21 one can notice that the alignment between $K_{A@Pia}$ and $K_{A@Rod}$ is greater than the alignment between $K_{A@Pia}$ and $K_{A@Sam}$. Thus, one can say that, even though Pia's understanding is not too similar to Rod's, her understanding is not opposite to his. This observation can be useful in situations where one wants to find an agreement on the meaning of a particular concept. For example, since their understandings are not completely opposite to each other, one can expect that Rod and Pia can reach an agreement in what should be understood as a Grandma's cookie more easily than Rod and Sam can do so.

Another option to compare Pia's, Rod's and Sam's understandings about Grandma's cookies is to use the *semantic differential method* proposed by Osgood in [10]. That method is based on bipolar scales denoted by polar terms like *interesting* and *boring* by which one can represent a perception or judgment.

In that regard, one can use polar terms such as *"'no-icing' dissatisfies the fulfillment of $\mathcal{C}$"* and *"'no-icing' satisfies the fulfillment of $\mathcal{C}$"* to represent

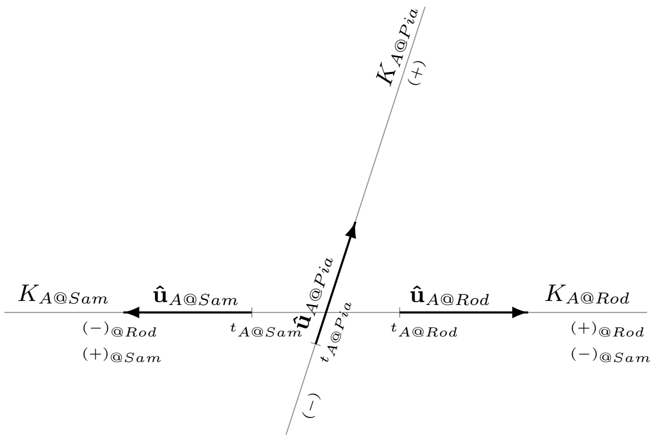


Figure 2.21: A graphical view of the difference in understandings about the Grandma's cookie concept.

the components of $\hat{\mathbf{u}}_{A@Pia}$, $\hat{\mathbf{u}}_{A@Rod}$ and $\hat{\mathbf{u}}_{A@Sam}$ in $K_{A@Pia}$, $K_{A@Rod}$ and $K_{A@Sam}$ respectively as shown in Figure 2.22. Notice in this figure that the influence of *'linear-icing'* in $K_{A@Rod}$ looks rather similar to the influence of this feature in $K_{A@Pia}$, which can be an indicator of how aligned Rod's and Pia's understandings are. However, one should also notice that, even though the training data set used during Sam's learning process includes completely opposite examples in relation to the training data set used during Rod's learning process, the influence of features like *'round-shape'* or *'square-shape'* seems to be the same according to what is depicted in the figure – in this case, someone can be misled into thinking that Rod's understanding is more similar to Sam's than Pia's.

The implications of the above observations for the design of methods by which one can compare XBEs will be described in the next chapters. Notwithstanding, we can anticipate that, as has been shown throughout this chapter, any significative difference between Pia's and Rod's understandings will mainly depend on the difference among the specific influence of the features deemed to be relevant by each of them.

2.5 Conclusions

In this chapter we have described a novel interpretation on how a person can experience a concept and how this experience is then reflected in his/her XBEs related to this concept. It was shown through a *feature-influence representational model* based on this interpretation that the knowledge acquired by a person after following a learning process with a particular training data set has

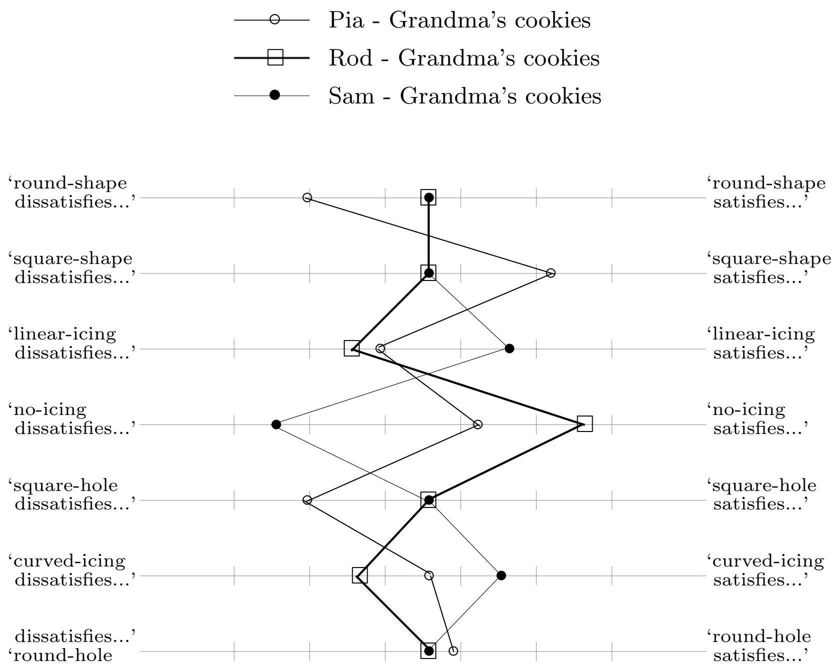


Figure 2.22: A semantic differential representation of Rod's, Pia's and Sam's understandings about Grandma's cookies.

an important influence on the context of his/her XBEs.

Since one can take into account the context of an XBE to assess how good (or bad) this XBE is, the above interpretation provides significant insights on how XBEs should be characterized to perform reliable comparisons among them. For instance, comparing XBEs having hints about what aspects have influenced the evaluation can be more reliable than comparing XBEs without such hints. In this regard, this interpretation gives practical guidelines on what a representation of an XBE should have and, thus, it helps to answer Research Question $Q1$ of this dissertation (see Section 1.5).

It was also shown that people following the same learning process with training data sets including opposite examples might provide totally opposite XBEs even though they focus on the same features. Since two respondents might have contrasting points of view during an evaluation process, one should be aware of this observation when comparing their XBEs. For instance, if someone is designing a comparison procedure, he/she should be aware that a comparison between the collections of XBEs given by two respondents with opposite understandings must result in the lowest degree of similarity. Hence, this observation can help to verify if a method proposed to compare XBEs given by a heterogeneous group of respondents is reliable or not – i.e., this observation can help to address Research Question $Q2$. It can also help to estimate the level to which those respondents share a similar understanding of the topic under analysis and, thus, identify which XBEs are given by respondents with whom a requester shares a similar understanding – i.e., it can also help to answer Research Questions $Q3$ and $Q4$.

This interpretation will be used in the next chapter to study how to handle XBEs with intuitionistic fuzzy sets (IFSs). It will also be used in Chapters 4 and 5 to determine the level to which comparison methods defined in the IFS framework are suitable for use in comparison of XBEs.

References

[1] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern classification.* John Wiley & Sons, 2012.

[2] Gösta Ekman. *A direct method for multidimensional ratio scaling.* Psychometrika, 28(1):33–41, 1963.

[3] Vladimir N. Vapnik. *The nature of statistical learning theory.* Springer-Verlag New York, Inc., 1995.

[4] Vladimir N. Vapnik and Vlamimir Vapnik. *Statistical learning theory*, volume 1. Wiley New York, 1998.

[5] Corinna Cortes and Vladimir Vapnik. *Support-vector networks.* Machine Learning, 20(3):273–297, 1995.

[6] Edgar Osuna, Robert Freund, and Federico Girosi. *Training support vector machines: an application to face detection.* In Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, pages 130–136. IEEE, 1997.

[7] Thorsten Joachims. *Text categorization with support vector machines: Learning with many relevant features.* Springer, 1998.

[8] Christopher J. Burges. *A tutorial on support vector machines for pattern recognition.* Data mining and knowledge discovery, 2(2):121–167, 1998.

[9] Thorsten Joachims. *Making large-Scale SVM Learning Practical.* In B. Schölkopf, C. Burges, and A. Smola, editors, Advances in Kernel Methods - Support Vector Learning, chapter 11, pages 169–184. MIT Press, Cambridge, MA, 1999.

[10] Charles E. Osgood. *The nature and measurement of meaning.* Psychological bulletin, 49(3):197, 1952.

Chapter 3

Handling Experience-Based Evaluations with Intuitionistic Fuzzy Sets

Abstract

In Chapter 2, we studied an interpretation on how a person can experience a concept and how that experience can be reflected in forthcoming evaluations related to that concept. In this chapter, we use that interpretation to model experience-based evaluations (XBEs) as elements of an intuitionistic fuzzy set (IFS). We use such a model to estimate, through comparisons among the XBEs given by two persons, the level to which the individual understandings acquired by these persons about the concept under evaluation are alike (or different). In this regard, hypothesizing that a difference in understandings could be marked by a difference in the evaluations of one or more relevant objects, we propose the concept *connotation-differential print* (CDP). A CDP allows for representing such a difference in a form that makes itself available to computation. We illustrate how a CDP can be used within a comparison between two intuitionistic fuzzy sets characterizing XBEs to reflect in a better way a perceived similarity between them.

This chapter is a compilation of the following publications:

- Marcelo Loor and Guy De Tré. *Connotation-Differential Prints - Comparing What Is Connoted Through (Fuzzy) Evaluations. Proceedings of the International Conference on Fuzzy Computation Theory and Applications - Volume 1: FCTA, (IJCCI 2014)*, 127-136. Rome, Italy, 2014.
- Marcelo Loor and Guy De Tré. *Vector Based Similarity Measure for Intuitionistic Fuzzy Sets. Modern approaches in fuzzy sets, intuitionistic fuzzy sets, generalized nets and related topics. Volume I: Foundations* edited by K. Atanassov, M. Baczynski, J. Drewniak, J. Kacprzyk, M. Krawczak, E. Szmidt, M. Wygralak and S. Zadrozny, 125-142, SRI-PAS, 2014.

3.1 Introduction

In Chapter 2 we described a situation in which, after trying individually some cookies, two cousins learned about what a Grandma's cookie is and, then, they used that knowledge to evaluate other cookies. Therein we explained how different experiences about Grandma's cookies can lead to different understandings of that concept. We have also shown how a difference in understandings could be perceived by looking at the (feature-influence representational) models of such understandings.

In this chapter, we consider a situation in which one tries to estimate the level to which the understandings acquired by those cousins are alike (or different) by comparing the XBEs given by them. This means that, in this case, we try to determine if a difference in understandings could be perceived by comparing the XBEs given by two persons. Our motivation here is to detect a kind of problem, called *pseudo-matching*, in which a comparison between two XBEs given by two persons can "match" even though these persons have different understandings of the evaluated concept.

With that end, we study if and how one can estimate such a difference by means of the framework of intuitionistic fuzzy sets (IFS) [1, 2]. In this framework, the evaluation of a proposition like "*cookie* 3 is a Grandma's cookie" can be expressed even if the evaluator is not fully convinced about its truthfulness. Hence, XBEs given by an evaluator who hesitates about his/her answers can be described more accurately in this framework.

To describe how to use the framework of IFS for handling XBEs, in the next section we explain how a collection of XBEs can be represented by an IFS. Then, we hypothesize that a difference in understandings could be marked by a difference in one or more evaluations and propose a novel technique by which a kind of footprint, called *connotation differential print* (CDP), is used to hint how different such understandings might be. After that, we explain how a CDP can be used to augment the results of similarity measures designed to compare IFSs.

3.2 Modeling XBEs as elements of IFSs

In the framework of *fuzzy set theory* [3], the evaluation of a proposition p having a canonical form 'x *IS* A' meaning "the value of (a subject) x is compatible with the definition of (a concept) A" [4] can be expressed in terms of a *membership grade*, which is a real number in the *unit interval* $[0, 1]$ that indicates the extent to which x is compatible with (or belongs to) A. In this regard, the evaluation of p for each element in a collection X can be represented by a *fuzzy set*, say A, which is mathematically denoted by

$$A = \{\langle x, \mu_A(x)\rangle | (x \in X) \wedge (0 < \mu_A(x) \leq 1)\}. \tag{3.1}$$

In [1, 2], Atanassov presented some examples in which, during the evaluation of p, it is better to indicate not only the extent to which x is compatible with (or belongs to) A, but also the extent to which x is *incompatible with* (or does not belong to) A. Thereby, he proposed an *intuitionistic fuzzy set* (IFS) as an extension of a fuzzy set.

An IFS, say A, is defined by

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | (x \in X) \wedge (0 \leq \mu_A(x) + \nu_A(x) \leq 1)\} \tag{3.2}$$

where the collection X is considered to be fixed, A represents the (intuitionistic fuzzy) subset of X under consideration, and the functions $\mu_A : X \to [0,1]$ and $\nu_A : X \to [0,1]$ define the *degree of membership* and the *degree of non-membership* of x in A respectively. In addition, the lack of knowledge about the membership (or non-membership) of x in A is expressed by

$$h_A(x) = 1 - \mu_A(x) - \nu_A(x) \tag{3.3}$$

and it is defined as the *degree of non-determinacy* – also known as *hesitation margin* [5].

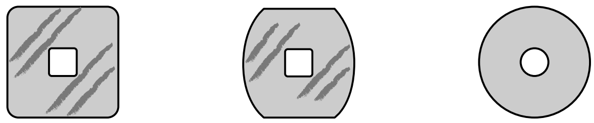
It is worth mentioning that, since the hesitation margin in fuzzy sets is implicitly assumed to be zero (i.e., there is no hesitation about the membership of x in A), the non-membership degree is expressed by the complement of the membership degree, i.e., $\nu_A(x) = 1 - \mu_A(x)$.

3.2.1 Semantic Interpretation of an IFS

IFSs offer excellent facilities to handle XBEs. To illustrate this, let us consider as a running example the following variant of the Grandma's cookies example presented in Chapter 2:

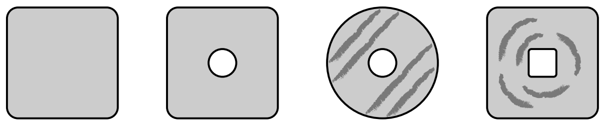
Example 3.1
Three cousins, Alice, Bob and Chloe, are individually evaluating to which degree a cookie could be seen or not as a Grandma's cookie. Each cousin has a mental picture of how looks a Grandma's cookie (see Figure 3.1), which is used as a frame of reference to evaluate all the cookies depicted in Figure 3.2. Using a unit interval scale where 1 *represents the highest level and* 0 *the lowest, the cousins have given their evaluations as shown in Table 3.1. It can be seen from the data in this table that, in some evaluations, adding both the 'yes'-value and the 'no'-value is not necessarily equal to* 1*. However, recording evaluations in this fashion allows the cousins to express any hesitation about their judgments.* Using the evaluations given by two cousins, how can someone estimate the level to which the mental pictures of the cookies used by these cousins are alike?

Using a semantic interpretation of the definition of an IFS, we can model the components of the above example as follows. The collection of all the cookies depicted in Figure 3.2 corresponds to the collection $X = \{$*cookie* 1, *cookie* 2, *cookie* 3, *cookie* 4$\}$. The level to which a cookie in X, say x, is deemed to be *compatible* with the way in which a Grandmas's cookie is perceived can be denoted by $\mu_A(x)$. In a similar way, the level to which a cookie in $x \in X$ is deemed to be *incompatible* with the way in which a Grandmas's cookie is perceived can be denoted by $\nu_A(x)$.



(a) *cookie a* (Alice). (b) *cookie b* (Bob). (c) *cookie c* (Chloe).

Figure 3.1: How looks a Grandma's cookie according to each relative (Grandma's cookie example).



(a) *cookie* 1. (b) *cookie* 2. (c) *cookie* 3. (d) *cookie* 4.

Figure 3.2: Do these cookies look like a Grandma's cookie? (Grandma's cookie example).

Using the degree-of-similarity semantic interpretation of a membership grade presented in [6], $\mu_A(x)$ can also be interpreted as the *degree of similarity* between x and the mental picture of a Grandma's cookies that a cousin may have. For instance, $\mu_{A@Alice}(cookie\ 1) = 0.6$ denotes the level to which Alice considers *cookie* 1 (see Figure 3.2a) to be similar to the way in which she perceives a Grandma's cookie (Figure 3.1a). Likewise, $\nu_{A@Alice}(cookie\ 1) = 0.3$ denotes the level to which *cookie* 1 is deemed to be dissimilar to the way in which a Grandma's cookie is perceived by Alice. In this case, the level to which Alice hesitates about considering *cookie* 1 to be similar or dissimilar to her mental picture of a Grandma's cookie can be denoted by $h_{A@Alice}(cookie\ 1) = 0.1$ and is obtained by $h_{A@Alice}(cookie\ 1) = 1 - (0.3 + 0.6)$.

The above interpretation means that the degrees of membership and non-membership can be obtained by following a process called *'membership exemplification'* [7]. In such a process, each cousin can provide a direct answer to the question '*to which degree a cookie is a Grandma's cookie?'* by using, e.g., a visual scale like the one shown in Figure 1.2 [8]. In this case, it is assumed

Table 3.1: Degree to which each cookie in Figure 3.2 can be seen as a Grandma's cookie by each cousin (Grandma's cookie example).

cookie	*yes*	*no*
cookie 1	0.6	0.3
cookie 2	0.7	0.3
cookie 3	0.2	0.8
cookie 4	0.9	0.1

(a) Alice.

cookie	*yes*	*no*
cookie 1	0.4	0.3
cookie 2	0.5	0.3
cookie 3	0	0.9
cookie 4	0.7	0.2

(b) Bob.

cookie	*yes*	*no*
cookie 1	0.6	0.3
cookie 2	0.7	0.3
cookie 3	0.2	0.8
cookie 4	0.1	0.9

(c) Chloe.

that a cookie is seen as a collection of *features* or *attributes*, which are used by each cousin to establish the degree of membership (or nonmembership) as follows [9]:

1. The degree of similarity (and, thus, the degree of membership) of a cookie will increase if the number of *common features* between this cookie and his/her mental picture of a Grandma's cookie increases. Analogously, the degree of dissimilarity (and, thus, the degree of nonmembership) of a cookie will increase if the number of *common features* between this cookie and his/her mental picture of a Grandma's cookie decreases.

2. The degree of similarity (and, thus, the degree of membership) of a cookie will decrease if the number of *distinctive features* between this cookie and his/her mental picture of a Grandma's cookie increases. Analogously, the degree of dissimilarity (and, thus, the degree of nonmembership) of a cookie will increase if the number of *distinctive features* between this cookie and his/her mental picture of a Grandma's cookie increases.

Using the above interpretation, the XBE of a cookie x can be modeled by an IFS element $\langle x, \mu_A(x), \nu_A(x)\rangle$. Hence, e.g., the XBEs given by Alice (see Table 3.1a can be represented by an IFS, say $A_{@Alice}$, such that

$$\begin{aligned} A_{@Alice} = \{&\langle cookie\ 1, 0.6, 0.3\rangle, \langle cookie\ 2, 0.7, 0.3\rangle, \\ &\langle cookie\ 3, 0.2, 0.8\rangle, \langle cookie\ 4, 0.9, 0.1\rangle\}. \end{aligned}$$

3.3 Detecting (Dis)similar Understandings

In Chapter 2, the notation $K_{A@P}$ was introduced to denote the knowledge (or understanding) acquired by a person P after following a learning process about (a concept) A. Using that notation in the running example of this chapter, we can say that $K_{A@Alice}$ and $K_{A@Bob}$ denote the understandings acquired by Alice and Bob respectively. In other words, $K_{A@Alice}$ and $K_{A@Bob}$ are mathematical representations of the mental pictures depicted in Figure 3.1a and Figure 3.1b respectively.

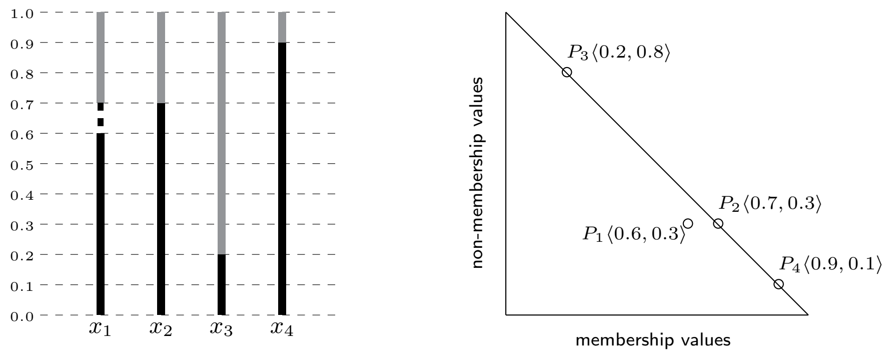
In that regard, the question *"how to estimate the level to which $K_{A@Alice}$ and $K_{A@Bob}$ are alike (or different) by means of a comparison of (the elements in) the IFSs $A_{@Alice}$ and $A_{@Bob}$?"* is raised in this section.

To find an answer to this question, we consider that a difference in understandings could be marked by a difference in one or more XBEs of *relevant objects* – here, by '*relevant object*' is meant an object having features that make the object a good example of compatibility (or incompatibility) with the concept under evaluation (e.g., *cookie* 4 can be deemed to be a relevant cookie by Alice since this cookie has features that make it compatible with the way in which she perceives a Grandma's cookie). This means that a difference between $K_{A@Alice}$ and $K_{A@Bob}$ could be determined through the comparisons of the IFS elements in $A_{@Alice}$ and $A_{@Bob}$ characterizing XBEs of one or more relevant cookies detected by Alice (or Bob).

In what follows we describe some geometric characteristics of the components of an IFS element and explain how a difference between two of them can be computed. After that, using the differences among the IFS elements representing XBEs of relevant objects, we describe how to build a kind of footprint designed to reflect any difference in the understandings that two persons may have regarding the concept under evaluation.

3.3.1 Geometric Characteristics of an IFS element

An IFS element has several geometric representations [2]. One of them is the mapping of the degrees of membership, non-membership and the hesitation margin of each element to an *unit segment*. For instance, Figure 3.3a depicts Alice's XBEs (see Table 3.1a) using this representation. In this figure, x_i represents a *cookie* i in X, the black-solid part of each unit segment denotes the degree of membership, the gray-solid part denotes the degree of non-membership, and the black-dotted one denotes the hesitation margin, respectively.



(a) Unit segments. (b) IFS-interpretational triangle.

Figure 3.3: Geometrical interpretations of Alice's XBEs.

Another geometric representation given in [2] is the so called *IFS-interpretational triangle*, in which the degrees of membership and non-membership of x are coordinates of a point P. For instance, in this interpretation Alice's XBEs can be represented as is shown in Figure 3.3b.

Rather than considering the degrees of membership and non-membership as coordinates of a point into the IFS-interpretational triangle, one can represent these components as a vector in $[0, 1]^2$, say $\mathbf{a}$, such that

$$\mathbf{a} = \begin{pmatrix} \mu_A(x) & + & \alpha_A \cdot h_A(x) \\ \nu_A(x) & + & (1 - \alpha_A) \cdot h_A(x) \end{pmatrix}, \tag{3.4}$$

where $\alpha_A \in [0, 1]$ is considered to be a *hesitation splitter* [10].

A hesitation splitter splits any hesitation about the compatibility (or incompatibility) of x in A. For instance, in Figure 3.4 the hesitation about the compatibility (or incompatibility) of x in A, which is denoted by $h_A(x)$, is split into the membership and the nonmembership components: while the α_A part of the hesitation is added to the membership component, the $(1 - \alpha_A)$ part is added to the nonmembership component. As can be noticed, a hesitation splitter allows us to distribute any hesitation between the membership or nonmembership components, which can be useful when a particular comparison strategy is needed.

A hesitation splitter semantically indicates in which proportion any hesitation about the compatibility of x in A will favor such compatibility. For instance, when $\alpha_A = 1$ holds, it means "any hesitation will entirely favor the compatibility of x in A." As such, a hesitation splitter can be fixed to reflect a particular comparison strategy: while in a *'pro membership strategy'* the value assigned to α_A is close to 1, in a *'pro non-membership strategy'* this value is close to 0.

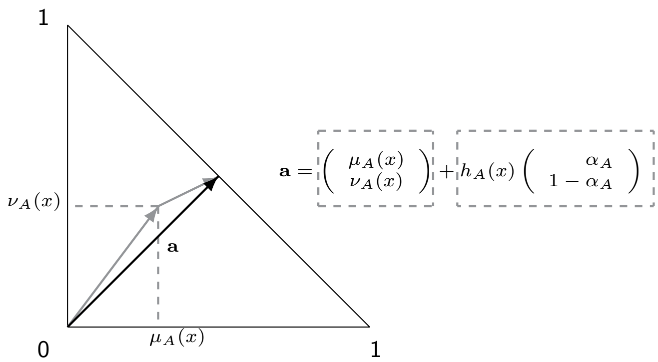


Figure 3.4: Vector representation of an IFS element.

Although they have different intentions, the hesitation splitter is somehow similar to the *extended modal operator* D_α, which is defined in [2] for an IFS A as

$$D_\alpha(A) = \{\langle x, \mu_A(x) + \alpha.h_A(x), \nu_A(x) + (1 - \alpha).h_A(x)\rangle | x \in X\}. \tag{3.5}$$

Consequently, taking as reference the extended modal operator $F_{\alpha,\beta}$ [2], which is defined by

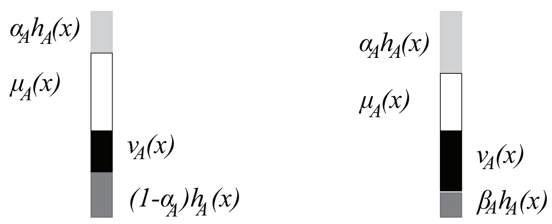
$$F_{\alpha,\beta}(A) = \{\langle x, \mu_A(x) + \alpha.h_A(x), \nu_A(x) + \beta.h_A(x)\rangle | x \in X\}, \tag{3.6}$$

it is possible to consider α_A and β_A as splitters such that $\alpha_A + \beta_A \leq 1$ where $\alpha_A, \beta_A \in [0, 1]$ (see Figure 3.5b). We will call α_A a *membership hesitation splitter* in A, and β_A a *non-membership hesitation splitter* in A. Thus, vector

$\mathbf{a}$ could also be expressed in terms of α_A and β_A as

$$\mathbf{a} = \begin{pmatrix} \mu_A(x) & + & \alpha_A \cdot h_A(x) \\ \nu_A(x) & + & \beta_A \cdot h_A(x) \end{pmatrix}. \tag{3.7}$$

In this case, while α_A semantically indicates in which proportion any hesitation about the compatibility of x in A will favor such compatibility, β_A indicates in which proportion such hesitation will disfavor such compatibility. These splitters can also be fixed to reflect a particular comparison strategy: when $\alpha_A > \beta_A$, it will be a '*pro membership strategy*,' when $\alpha_A < \beta_A$, it will be '*pro non-membership strategy*,' and when $\alpha_A = \beta_A$ the hesitation will not favor nor disfavor the compatibility of x in A.



(a) (b)

Figure 3.5: Unit-segment interpretations of hesitation splitters.

3.3.2 Spot Comparison

One can use the above *vector representation* of an IFS element to compare, e.g., the XBEs of a cookie given by Alice and Bob. To do so, considering $\mathbf{a}$ and $\mathbf{b}$ to be vectors representations of the IFS elements $\langle x, \mu_A(x), \nu_A(x)\rangle$ and $\langle x, \mu_B(x), \nu_B(x)\rangle$ that represent the XBEs of a cookie x given by Alice and Bob respectively, one can use a measure of the difference between $\mathbf{a}$ and $\mathbf{b}$, say $dif(\mathbf{a}, \mathbf{b})$, to estimate how different these XBEs are.

To compute such a *spot difference*, i.e., a difference between two specific IFS elements, one can use the area of the parallelogram formed by $\mathbf{a}$ and $\mathbf{b}$ as an indicator of the difference: the larger this area, the larger the difference between $\mathbf{a}$ and $\mathbf{b}$. This means that, within this approach, $dif(\mathbf{a}, \mathbf{b})$ will be related to the vector product between $\mathbf{a}$ and $\mathbf{b}$, i.e., $\mathbf{a} \times \mathbf{b}$. For instance, using the vector interpretation given in (3.4) one can obtain the expression

$$dif(\mathbf{a}, \mathbf{b}) = (\mu_A(x) - \mu_B(x)) + (\alpha_A \cdot h_A(x) - \alpha_B \cdot h_B(x)) \tag{3.8}$$

to compute the spot difference between $\mathbf{a}$ and $\mathbf{b}$. In a similar way, using the interpretation given in (3.7), one can obtain the expression

$$dif(\mathbf{a},\mathbf{b}) = (\mu_A(x) + \alpha_A \cdot h_A(x)) \cdot (\nu_B(x) + \beta_B \cdot h_B(x)) \\ - (\mu_B(x) + \alpha_B \cdot h_B(x_i)) \cdot (\nu_A(x) + \beta_A \cdot h_A(x)). \tag{3.9}$$

In the context of the running example, one can semantically interpret (3.8) as follows. The first part, i.e., $(\mu_A\,(x) - \mu_B\,(x))$, denotes that the spot difference between Alice's and Bob's evaluations of x is determined partly by the difference between the levels to which x is deemed to be a Grandma's cookie by each of them. The second part, i.e., $(\alpha_A \cdot h_A(x)\ {-\alpha_B} \cdot h_B(x))$, denotes that the spot difference is also influenced by any doubt about considering x to be a Grandma's cookie. The sign $(+/-)$ of (3.8) denotes the *relative difference* between Alice and Bob's evaluations: when $dif(\mathbf{a},\mathbf{b}) > 0$, it means that Alice considers x to be *more* compatible with a Grandma's cookie than Bob considers so; and when $dif(\mathbf{a},\mathbf{b}) < 0$, it means that Alice considers x to be *less* compatible with a Grandma's cookie than Bob considers so.

Since both (3.8) and (3.9) can be affected by both Alice's and Bob's hesitation splitters, i.e., α_A, α_B, β_A and β_B, one can establish a comparison strategy in which the same rule is applied for these hesitation splitters. By doing so, $\alpha_A = \alpha_B = \alpha$ and $\beta_A = \beta_B = \beta$ will hold and, thus, (3.8) and (3.9) can be rewritten as

$$dif^{\alpha}(\mathbf{a},\mathbf{b}) = (\mu_A\,(x) - \mu_B\,(x)) + \alpha\,(h_A(x) - h_B(x)) \tag{3.10}$$

and

$$dif^{\alpha,\beta}(\mathbf{a},\mathbf{b}) = (\mu_A(x) + \alpha \cdot h_A(x)) \cdot (\nu_B(x) + \beta \cdot h_B(x)) \\ - (\mu_B(x) + \alpha \cdot h_B(x_i)) \cdot (\nu_A(x) + \beta \cdot h_A(x)) \tag{3.11}$$

respectively.

Another approach to compute a spot difference is one in which the distance between **a** and **b** is used as an indicator of the difference: the larger the distance, the larger the difference between **a** and **b**. For instance, using the vector representation given in (3.7) within this approach, one can obtain the expression

$$dif(\mathbf{a},\mathbf{b}) = \big(((\mu_A(x) - \mu_B(x)) + (\alpha_A \cdot h_A(x) - \alpha_B \cdot h_B(x)))^2 \\ + ((\nu_A(x) - \nu_B(x)) + (\beta_A \cdot h_A(x) - \beta_B \cdot h_B(x)))^2\big)^{\frac{1}{2}}. \tag{3.12}$$

In this case, assuming that $\alpha_A = \alpha_B = \alpha$ and $\beta_A = \beta_B = \beta$ holds, one can rewrite (3.12) as

$$dif(\mathbf{a},\mathbf{b}) = \big(((\mu_A(x) - \mu_B(x)) + \alpha(h_A(x) - h_B(x)))^2 \\ + ((\nu_A(x) - \nu_B(x)) + \beta(h_A(x) - h_B(x)))^2\big)^{\frac{1}{2}}. \tag{3.13}$$

In a similar way to the semantic interpretation of (3.8), one can say that a spot difference computed by (3.12) or (3.13) depends on the levels of membership, non-membership and hesitation recorded in the IFS elements representing the XBEs given by Alice and Bob. However, in contrast to (3.8) and (3.9), a spot difference computed by (3.12) or (3.13) does not provide a (+/-) sign as an indicator of the relative difference between Alice's and Bob's XBEs. This means that, in this case, we could not determine by looking at a result computed by these equations if Alice considers x to be more compatible with a Grandma's cookie than Bob considers so. Hence, in the next part, we will study how to use spot differences computed by variants of (3.8) or (3.9) to determine how similar or dissimilar the understandings behind two XBEs are.

3.3.3 Connotation Differential Print

As was mentioned at the beginning of this section, one can consider that a difference in understandings could be marked by a difference in one or more XBEs of relevant objects. In this regard, one can use the spot differences of IFS elements related to XBEs of relevant objects to detect potential differences in understandings about the concept under evaluation.

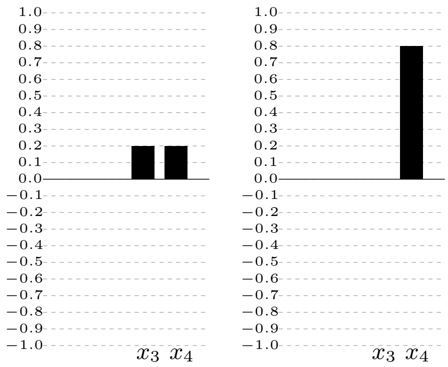
For instance, consider that *cookie* 3 and *cookie* 4 (see Figure 3.2) are the relevant cookies identified by Alice. Consider also that $\mathbf{a}_3$, $\mathbf{a}_4$, $\mathbf{b}_3$, $\mathbf{b}_4$, $\mathbf{c}_3$ and $\mathbf{c}_4$ are vector representations of the XBEs of *cookie* 3 and *cookie* 4 given by Alice, Bob and Chloe respectively (see Table 3.1). With these considerations in mind, one can use (3.10) with $\alpha = 0$ to obtain $dif(\mathbf{a}_3, \mathbf{b}_3) = 0.2$, $dif(\mathbf{a}_4, \mathbf{b}_4) = 0.2$, $dif(\mathbf{a}_3, \mathbf{c}_3) = 0$ and $dif(\mathbf{a}_4, \mathbf{c}_4) = 0.8$. Since *cookie* 3 and *cookie* 4 have some features that make them good examples of compatibility (or incompatibility) with a Grandma's cookie according to Alice, these results suggest that the difference between her understanding about a Grandma's cookie and Bob's is less than the difference between her understanding and Chloe's. In other words, the results indicate that the mental picture of a Grandma's cookie in Alice's mind is likely more similar to Bob's than Chloe's mental picture.

For a visualization of the above results, the spot differences can be depicted as shown in Figure 3.6a and Figure 3.6b. In these figures, each spot difference is represented by a ruler of height one, marked of with "difference"-units. The black region denotes the magnitude of a spot difference, and the position of the black region, above or below the line that represents no-difference, denotes its relative difference.

To illustrate how such spot differences can help to detect (dis)similar understandings, one can apply the set-theoretical approach of similarity proposed by Tversky in [9], in which the similarity between two objects is deemed to result from a *feature-matching* process.

Using that approach, each cookie in the *Grandma's cookie* example can be seen as an object with features such as a square shape, linear icing, or with a square hole. Thus, in the abstraction process carried out to give her XBEs, Alice could pay more attention than Bob or Chloe to some cookie's features according to her memory or mental picture of a Grandma's cookie, causing a difference in their individual understandings – recall the experience-based evaluation process described in Section 2.3.

The mental pictures of Grandma's cookies preserved by each cousin are



(a) Alice vs. Bob. (b) Alice vs. Chloe.

Figure 3.6: Visual representations of spot differences.

depicted in Figure 3.1. Notice that, while Alice's memory about a Grandma's cookie is a cookie with a square shape, linear icing and a square hole, Bob's and Chloe's memories are a cookie with a round square shape, linear icing and a square hole, and a cookie with round shape, no icing and a round hole respectively.

When Alice evaluated to which degree *cookie* 3 (see Figure 3.2c) can be seen as a Grandma's cookie, she judged it as 0.2 for 'yes' and 0.8 for 'no' (see Table 3.1a) – it seems that Alice paid attention to the round shape or the round hole of the cookie and not to linear icing, which is a common feature between *cookie* 3 and her memory of a Grandma's cookie. When Chloe did so, she also judged it as 0.2 for 'yes' and 0.8 for 'no' (see Table 3.1c) – although the round shape and the round hole are features present in both *cookie* 3 and Chloe's mental picture of a Grandma's cookie, it seems that she paid attention to the linear icing during her evaluation.

If Alice takes only into account the visual representation of the spot differences related to *cookie* 3, she might wrongly interpret that her understanding about a Grandma's cookie is more similar to Chloe's than Bob's. Nevertheless, looking at the visual representation of the spot differences related to *cookie* 4, Alice can notice that Chloe seems to pay less attention to the square shape or square hole in *cookie* 4. Hence, Alice can realize that her understanding about a Grandma's cookie differs from Chloe's. This observation suggests that, since the spot differences are built from the XBEs of relevant cookies detected by Alice, she needs to look at all of them to try to identify understandings similar or dissimilar to hers.

To make such spot differences available for computation and, thus, make an estimation of the level of (dis)similarity between the understandings of two persons, we propose the use of a kind of marker that reflects the relative difference between two XBEs characterized as IFS elements. Such a marker is defined as follows:

Definition 3.1 (Connotation Differential Marker)
Consider an element $x \in X$. Let **a** *be a vector representing the membership and non-membership of x to the IFS A,* **b** *a vector representing the membership and non-membership of x to the IFS B, and $dif(\mathbf{a}, \mathbf{b})$ a spot difference between* **a** *and* **b**. *Now consider a set $S = \{\varphi, \uparrow, \downarrow\}$. A* connotation differential marker*, CDM for short, is a symbol $s \in S$ that denotes the perceived difference between* **a** *and* **b** *according to the following conditions:*

- *if $|dif(\mathbf{a}, \mathbf{b})| \leq \delta$ then $s = \varphi$,*
- *if $dif(\mathbf{a}, \mathbf{b}) > \delta$ then $s = \uparrow$,*
- *if $dif(\mathbf{a}, \mathbf{b}) < -\delta$ then $s = \downarrow$,*

where $\delta \in [0, 1]$.

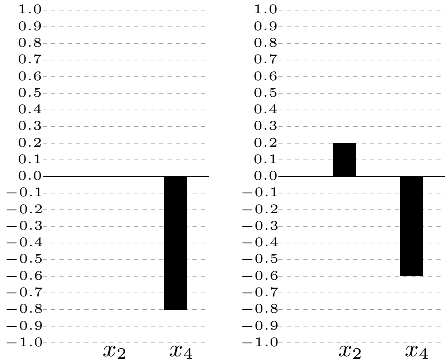
Using the above definition with $\delta = 0.2$, the spot difference between Alice and Bob for *cookie* 4, i.e., $dif(\mathbf{a}_4, \mathbf{b}_4) = 0.2$, can be denoted by φ. By contrast, the spot difference between Alice and Chloe for this cookie, i.e., $dif(\mathbf{a}_4, \mathbf{c}_4) = 0.8$, can be denoted by $\uparrow$.

Knowing what a CDM could denote individually, we can put one or more CDMs together in order to obtain a representation that hints if a cousin has paid attention or not to the same cookie's features that have been focused by another during the evaluation process. A way to do that is by placing one or more CDMs in a sequence with a particular order.

For example, to perform the comparisons from Alice's point of view, we can build sequences with two CDMs: the first one related to her best evaluated cookie, i.e., *cookie* 4 (see Table 3.1a), and the second one, to her worst evaluated cookie, i.e., *cookie* 3. Thus, with $\delta = 0.2$, while a sequence corresponding to (the comparison) Alice vs. Bob is "$\varphi\varphi$", the sequence corresponding to Alice vs. Chloe is "$\uparrow\varphi$".

Despite using only two CDMs, looking at the Alice-vs.-Bob sequence, Alice can distinguish that Bob seems to agree with her about *cookie* 4 having one or more features to consider it to be a Grandma's cookie, as well as, *cookie* 3 having one or more features to consider it not to be so. On the other hand, looking at the first CDM in the Alice-vs.-Chloe sequence, Alice can become aware that some features of *cookie* 4 make Chloe to consider it not to be a Grandma's cookie and, thus, Alice can realize that she and Chloe have different connotations of a Grandma's cookie. Because of this, we call such a sequence a *connotation differential print*, or CDP for short.

Now, let us perform the comparison from Chloe's point of view. Using the same strategy to build the above sequences, the first and the second CDM are related to the best and the worst evaluated cookies respectively, i.e. *cookie* 2 and *cookie* 4 (see Table 3.1c). Thus, the CDP corresponding to Chloe-vs.-Bob comparison (see Figure 3.7a) is "$\varphi\downarrow$", and the CDP corresponding to Chloe-vs.-Alice comparison (see Figure 3.7b) is "$\varphi\downarrow$". Looking at these CDPs, Chloe can realize that neither Bob nor Alice remembers a Grandma's cookie with no icing as she does.



(a) Chloe vs. Alice. (b) Chloe vs. Bob.

Figure 3.7: Visual representations of connotation differential prints.

As could be noticed above, a CDP depends on the individual point of view of each cousin. In fact, Alice has chosen the CDMs corresponding to *cookie* 4 and *cookie* 3, while Chloe has chosen *cookie* 2 and *cookie* 4. This is an example of directionality and asymmetry in comparison judgments pointed out by Tversky in [9].

To facilitate a comparison between two CDPs, Alice can assign a weight to each of the possible CDPs in order to determine which understanding is more similar to hers. For example, according to her strategy to build a CDP, ⟠⟠ denotes a good level of similarity, thus, she assigns 1.0 to it. The CDPs ⫯⫯ and ⫰⫰ denote a not too bad similarity (these CDPs could become ⟠⟠ by increasing δ), therefore she gives 0.75 to them. The CDPs ⫯⟠, ⟠⫯, ⫰⟠ and ⟠⫰ denote a big difference, so, she gives 0.25 to them. Finally, ⫯⫰ and ⫰⫯ denote a huge difference, so, she assigns 0 to them.

Recalling the question raised at the beginning of this section, we can say at this point that a CDP constitutes an estimation of the level to which the understanding acquired by someone (e.g., Alice) about a particular concept (e.g., Grandma's cookies) is similar (or different) to the understanding acquired by someone else (e.g., Bob or Chloe).

In the next section, we explain how a CDP can be used for comparing two IFSs representing XBEs to reflect in a better way a perceived similarity between them.

3.4 Comparing IFSs that represent XBEs

After characterizing a collection of XBEs as an IFS, one can theoretically use any of the existing similarity measures in the IFS framework to compare such collections. A similarity measure in this framework can be defined as follows:

Definition 3.2
Let A and B be two IFSs in $X = \{x_1, \cdots, x_n\}$. A similarity measure*, say S, is a mapping $S : X^2 \mapsto [0, 1]$ such that $S(A, B)$ denotes the level to which A is similar to B with 0 and 1 representing the lowest and the highest (similarity) levels respectively.*

In the IFS literature one can identify a kind of *symmetric* approach in the formulation of similarity measures. In this approach, the similarity between two IFSs is usually assumed to be a *"dual notion of a metric distance"* [5]. Thus, given a normalized metric distance function $d : X^2 \mapsto [0, 1]$, the similarity S between A and B can be expressed as $S(A, B) = 1-d(A, B)$, where d follows the axioms of *minimality* (i.e., $d(A, A) = 0$), *symmetry* (i.e., $d(A, B) = d(B, A)$) and the *triangle inequality* (i.e., $d(A, B) + d(B, C) \geq d(A, C)$). In this regard, a *symmetric similarity measure* can be defined as follows:

Definition 3.3
Let A and B be two IFSs in $X = \{x_1, \cdots, x_n\}$. A similarity measure, say S, is said to be symmetric*, if and only if $S(A, B) = S(B, A)$ holds for any A, B in X.*

The following are some examples of symmetric similarity measures:

$$
\begin{aligned}
S_{H3D}(A, B) = 1 - \frac{1}{2n} \sum_{i=1}^{n} (&|\mu_A(x_i) - \mu_B(x_i)| \\
&+ |\nu_A(x_i) - \nu_B(x_i)| \\
&+ |h_A(x_i) - h_B(x_i)|)
\end{aligned}
\tag{3.14}
$$

and

$$
\begin{aligned}
S_{H2D}(A, B) = 1 - \frac{1}{2n} \sum_{i=1}^{n} (&|\mu_A(x_i) - \mu_B(x_i)| \\
&+ |\nu_A(x_i) - \nu_B(x_i)|)
\end{aligned}
\tag{3.15}
$$

which are based on the Hamming distance with the hesitation component [11] and the Hamming distance without the hesitation component [2] respectively;

$$
\begin{aligned}
S_{E3D}(A, B) = 1 - \Bigg(\frac{1}{2n} \sum_{i=1}^{n} \big(&(\mu_A(x_i) - \mu_B(x_i))^2 \\
&+ (\nu_A(x_i) - \nu_B(x_i))^2 \\
&+ (h_A(x_i) - h_B(x_i))^2 \big) \Bigg)^{\frac{1}{2}}
\end{aligned}
\tag{3.16}
$$

and

$$S_{E2D}(A,B) = 1 - \left(\frac{1}{2n} \sum_{i=1}^{n} \left(\left(\mu_A(x_i) - \mu_B(x_i) \right)^2 \right. \right.$$
$$\left. \left. + \left(\nu_A(x_i) - \nu_B(x_i) \right)^2 \right) \right)^{\frac{1}{2}} \tag{3.17}$$

based on the Euclidean distance with the hesitation component [11] and the Euclidean distance without the hesitation component [2] respectively.

To compare IFSs characterizing XBEs, we prefer to use a psychologically driven approach in the formulation of similarity measures since an XBE depends of the experience of a person. In contrast to the symmetric approach, in this psychological approach, the similarity between the IFSs A and B is assumed to be the result of the evaluation of the proposition 'A is like B,' which is not the same as the proposition 'B is like A.' While in the former proposition, B is deemed to be the frame of reference, in the latter, A is deemed to be so. For instance, let A and B be two IFSs characterizing the XBEs given by Alice and Bob respectively. If A is taken as a frame of reference, Alice's mental picture of a Grandma's cookie will be taken into account to evaluate the similarity between A and B. By the contrary, Bob's mental picture of a Grandma's cookie will be considered when B is taken as a reference – because of this directionality, this approach is also called a *directional* approach.

Such directional similarity statements were studied by Tversky in [9]. In that work, Tversky provided empirical evidence suggesting that a similarity comparison is better described as a process in which someone compares the (collection of) features detected in two objects instead of a process in which someone computes a metric distance between those objects. Hence, a similarity measure based on the directional approach can have the form

$$S(A,B) = \lambda_1 f(A \cap B) - \lambda_2 f(A - B) - \lambda_3 f(B - A), \tag{3.18}$$

which is called *contrast model*, or it can have the form

$$S(A,B) = \frac{f(A \cap B)}{f(A \cap B) + \lambda_2 f(A - B) + \lambda_3 f(B - A)}, \tag{3.19}$$

which is called *ratio model*. In these equations, while λ_1, λ_2 and λ_3 are numbers greater or equal to 0, f is a non-negative measure of the contributions of $A \cap B$, $A - B$ or $B - A$. In this regard, a *directional similarity measure* can be defined as follows:

Definition 3.4
Let A and B be two IFSs in $X = \{x_1, \cdots, x_n\}$. A similarity measure, say S, is said to be directional, *if $S(A,B)$ is not necessarily equal to $S(B,A)$ for any A, B in X.*

An example of a similarity measure following the directional approach is given by the expression

$$S(A,B) = \frac{g_{A \cap B}(A,B)}{g_{A \cap B}(A,B) + \lambda_2 \cdot g_{A-B}(A,B) + \lambda_3 \cdot g_{B-A}(A,B)}, \tag{3.20}$$

where

$$g_{A-B}(A,B) = \begin{cases} \sum_{i=1}^{n} dif(\mathbf{a_i},\mathbf{b_i}) & (\forall i : dif(\mathbf{a_i},\mathbf{b_i}) > 0) \\ 0 & \text{otherwise}, \end{cases}$$

$$g_{B-A}(A,B) = \begin{cases} \sum_{i=1}^{n} |dif(\mathbf{a_i},\mathbf{b_i})| & (\forall i : dif(\mathbf{a_i},\mathbf{b_i}) < 0) \\ 0 & \text{otherwise}, \end{cases}$$

and

$$g_{A\cap B}(A,B) = n - g_{A-B}(A,B) - g_{B-A}(A,B).$$

With $\lambda_2 = \lambda_3 = 1$ and $dif(\mathbf{a_i},\mathbf{b_i}) = dif^{\alpha}(\mathbf{a_i},\mathbf{b_i})$ (see (3.10)), one can rewrite (3.20) as follows [10]:

$$S^{\alpha}(A,B) = 1 - \frac{1}{n}\sum_{i=1}^{n} |dif^{\alpha}(\mathbf{a_i},\mathbf{b_i})|. \tag{3.21}$$

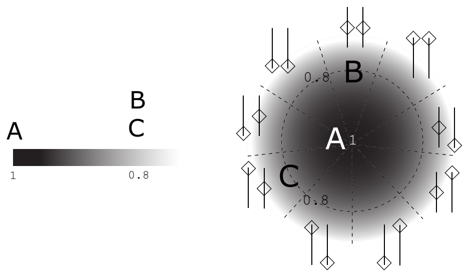
A main difference between the aforementioned approaches is that, while in a symmetric approach, the expression $S(A,B) = S(B,A)$ always holds, in a directional approach this expression only holds when (the understandings of) the concepts denoted by A and B are the same, i.e., $S(A,B) = S(B,A)$ only holds when the frames of reference (or mental pictures) to evaluate the compatibility of any $x_i \in X$ in A and B respectively are alike.

Although (3.21) takes into account a potential human behavior during a similarity comparison, it can only obtain the magnitude of the similarity between A and B. Therefore, one can say nothing about the correspondence between the appearance of the frames of reference (or mental pictures) used to evaluate the compatibility of any $x_i \in X$ in A and B.

For instance, considering A, B and C the IFSs that represent the XBEs given by Alice, Bob and Chloe respectively, one can compute with (3.21) the similarity between Alice and Bob and obtain $S^0(A,B) = 0.8$ as a result. Likewise, one can compute the similarity between Alice and Chloe and obtain $S^0(A,C) = 0.8$. Notice here that, since nothing is said about the correspondence between the mental picture of a Grandma's cookie used by the cousins, these results are the same even though Alice's mental picture looks more similar to the Bob's than the Chloe's mental picture (see Figure 3.1).

To achieve better fine-tuned and more reliable (similarity) comparisons, we propose using a CDP to *augment the results* produced by a similarity measure. Since a CDP can be used to hint how different such mental pictures might be, one can consider that, if a similarity measure tells us about *how far is A from B*, an *augmented similarity measure* with a CDP will additionally tell us *to which extend B is aligned with A*.

As an example, consider Figure 3.8, in which the comparisons *Alice-vs.-Bob* and *Alice-vs.-Chloe*, i.e., $S^0(A,B) = 0.8$ and $S^0(A,C) = 0.8$, are depicted:



(a) (b)

Figure 3.8: Alice vs. Bob and Alice vs. Chloe similarities.

in (a) only the magnitude is used and, so, there is no difference between both similarities; in (b) the magnitude plus a CDP are used and, so, it is noteworthy how the "direction" of *Alice-vs.-Bob* differs from *Alice-vs.-Chloe*'s. Notice that, from Alice's point of view, the augmented similarity between her evaluations and Bob's can be denoted by $\langle 0.8, \rangle$ while, in comparison to Chloe's evaluations, it can be denoted by $\langle 0.8, \rangle$. Moreover, using the weights that Alice proposed earlier, we could say that *Alice-vs.-Bob*'s $\langle 0.8, \rangle$ is $0.8 \cdot 1 = 0.8$, and *Alice-vs.-Chloe*'s $\langle 0.8, \rangle$ is $0.8 \cdot 0.25 = 0.2$. This reflects that Alice's XBEs are more similar to Bob's than to Chloe's.

3.4.1 Some difficulties with similarity measures

In [5], Szmidt and Kacprzyk have examined the effects of the assumption of symmetry in similarity measures designed to compare IFSs. They found some difficulties that, according to them, are a result of (i) the symmetry of the three components of a IFS element (i.e., the membership, non-membership and hesitation values) and (ii) the role played by these components in the definition of the complement of IFSs – which should be considered in such similarity measures.

One of the difficulties is exemplified as follows. Consider $X = \{x_0\}$ and the IFSs $M = \{\langle x_0, 1, 0\rangle\}$, $N = \{\langle x_0, 0, 1\rangle\}$ and $H = \{\langle x_0, 0, 0\rangle\}$. Also consider the similarity measure S. If S is (3.14) or (3.16), it is obtained that $S(M, N) = S(M, H)$ even though N and H are different – if this result is used, e.g., in a clustering process, N and H might be wrongly included in the same group. This anomaly is generalized to IFSs such as $K = \{\langle x_0, 0.5, 0.3\rangle\}$ and $L = \{\langle x_0, 0.5, 0.2\rangle\}$ where the exchange of "the places" between the non-membership value and the hesitation margin in K and L results in $S(M, K) = S(M, L)$. Since this anomaly is caused by the symmetry between the non-membership value and the hesitation margin, Szmidt and Kacprzyk also verified the results using the "two terms"-distances (3.15) and (3.17). However, they found that the situation does not change in the sense of the information obtained.

As a solution to this kind of anomalies, we proposed (3.21) in [10]. Thus, in this example, $S^\alpha(M, N) = 0$ and $S^\alpha(M, H) = \alpha$, which makes sense according to the semantic interpretation given in Section 3.3.2. Despite this, there is an anomaly that (3.21) could not manage. So, a CDP is needed.

Consider $X = \{x_0\}$ and the IFSs $P = \{\langle x_0, 0.5, 0.5\rangle\}$, $Q = \{\langle x_0, 0.9, 0.1\rangle\}$ and $R = \{\langle x_0, 0.1, 0.9\rangle\}$. Also consider the similarity measure S. If S is (3.14), (3.15), (3.16), (3.17) or even (3.21), it is obtained that $S(P, Q) = S(P, R)$ even though Q and R are obviously different. This anomaly could be generalized to IFSs such as $V = \{\langle x_0, 0.7, 0.3\rangle\}$ and $W = \{\langle x_0, 0.3, 0.7\rangle\}$ where giving the values $\mu_V(x_0) = \nu_W(x_0)$, $\mu_W(x_0) = \nu_V(x_0)$ and $h_V(x_0) = h_W(x_0) = 0$ results in $S(P, V) = S(P, W)$.

To solve the above anomaly, we use the augmented version of (3.21) as follows. From (3.4) the corresponding vector interpretations for x_0 are $\mathbf{p_0} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$, $\mathbf{q_0} = \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix}$ and $\mathbf{r_0} = \begin{pmatrix} 0.1 \\ 0.9 \end{pmatrix}$ – notice that $h_P(x_0) = 0$, $h_Q(x_0) = 0$ and $h_R(x_0) = 0$, which means that a hesitation splitter is not necessary. Then, considering the point of view of P and using (3.10), the spot-differences for x_0 are $dif^{\alpha}(\mathbf{p_0}, \mathbf{q_0}) = -0.4$ and $dif^{\alpha}(\mathbf{p_0}, \mathbf{r_0}) = 0.4$. With $\delta = 0.2$, the corresponding connotation-differential marker for $dif^{\alpha}(\mathbf{p_0}, \mathbf{q_0})$ is $\underset{\diamond}{\downarrow}$ and the corresponding one for $dif^{\alpha}(\mathbf{p_0}, \mathbf{r_0})$ is $\overset{\diamond}{\uparrow}$. To build the CDPs for P-vs.-Q and P-vs.-R comparisons, we use the connotation-differential markers given for x_0, thus, from P's view, $\underset{\diamond}{\downarrow}$ is a CDP for P-vs.-Q, and $\overset{\diamond}{\uparrow}$ is a CDP for P-vs.-R. Finally, from (3.21) we obtain $S^{\alpha}(P, Q) = 0.6$ and $S^{\alpha}(P, R) = 0.6$, and, using the corresponding CDPs, we augment them to $\langle 0.6, \underset{\diamond}{\downarrow}\rangle$ and $\langle 0.6, \overset{\diamond}{\uparrow}\rangle$ respectively. As expected, $\langle 0.6, \underset{\diamond}{\downarrow}\rangle$ and $\langle 0.6, \overset{\diamond}{\uparrow}\rangle$ are different.

One might argue that if the same weight is assigned to $\underset{\diamond}{\downarrow}$ and $\overset{\diamond}{\uparrow}$, then $S(P, Q) = S(P, R)$. This illustrates that weight assignment is an important and delicate task as it should adequately reflect the semantics of the CDPs.

3.5 Related Work

About the semantic interpretation of the elements of an IFS (see Section 3.2), we found in the theoretical model proposed by Ekman in [12] – which is used to compare two perceptions from an observer – some analogies that fit with those used in our vector based interpretation of the membership and non-membership of such elements. In his model, Ekman considered that, while the perceptual intensity can be depicted by the magnitude of a vector, the perceptual quality is given by the vector's direction. Analogically, the perceptual intensity corresponds to the degree to which one element, say x, belongs or not to an IFS A, i.e., $\mu_A(x)$ and $\nu_A(x)$ respectively; while the perceptual quality corresponds to agreement on what is understood by A.

3.6 Conclusions

In this chapter, we explained how to model a collection of experience-based evaluations (XBEs) as an intuitionistic fuzzy set (IFS) and, also, how to use such IFSs to compare XBEs.

Although an IFS can only represent the levels but not the aspects that might influence the XBEs, it becomes an option to handle subjective, imprecise and potentially marked-by-hesitation XBEs that do not include hints about such aspects. In this regard, IFSs can be deemed to be an option to answer Research Question $Q1$.

Since the aspects that might influence the XBEs could not be recorded in an IFS, we described a novel approach to compare them. In this approach, the XBEs of some relevant objects are used to build a kind of footprint of the comparison. This footprint, named *connotation-differential print* (CDP), constitutes a representation of a possible difference in the understandings that two persons might have about the topic under evaluation.

An illustrative example presented in this chapter shows that a CDP is suitable for computation and, as such, can be used to augment the results of similarity measures in order to achieve more reliable (similarity) comparisons among IFSs characterizing XBEs.

In that regard, a CDP can help to perform a simple but acceptable comparison between XBEs given by a heterogeneous group of respondents (Research Question $Q2$). This means that a CDP can be used to detect XBEs given by respondents with whom a requester shares a similar understanding (Research Question $Q4$) and, thus, compute an estimation of the quality perceived by this requester on such XBEs (Research Question $Q3$).

To complement the theoretical study presented in this chapter, an empirical process by which several similarity measures are tested while comparing IFSs that represent XBEs will be described in the next chapter.

References

[1] Krassimir T. Atanassov. *Intuitionistic fuzzy sets.* Fuzzy sets and Systems, 20(1):87–96, 1986.

[2] Krassimir T. Atanassov. *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

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

[4] Lotfi A. Zadeh. *Is there a need for fuzzy logic?* Information Sciences, 178(13):2751 – 2779, 2008.

[5] Eulalia Szmidt and Janusz Kacprzyk. *Geometric similarity measures for the intuitionistic fuzzy sets.* In 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13). Atlantis Press, 2013.

[6] Didier Dubois, Walenty Ostasiewicz, and Henri Prade. *Fuzzy sets: history and basic notions.* In Fundamentals of Fuzzy Sets, pages 21–124. Springer, 2000.

[7] Taner Bilgiç and I Burhan Türkşen. *Measurement of membership functions: theoretical and empirical work.* In Fundamentals of fuzzy sets, pages 195–227. Springer, 2000.

[8] Jean-Lou Chameau and Juan Carlos Santamarina. *Membership functions I: Comparing methods of measurement.* International Journal of Approximate Reasoning, 1(3):287 – 301, 1987.

[9] Amos Tversky. *Features of similarity.* Psychological review, 84(4):327, 1977.

[10] Marcelo Loor and Guy De Tré. *Vector based similarity measure for intuitionistic fuzzy sets.* In New Trends in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics, 2013.

[11] Eulalia Szmidt and Janusz Kacprzyk. *Distances between intuitionistic fuzzy sets.* Fuzzy Sets and Systems, 114(3):505–518, September 2000.

[12] Gösta Ekman. *A direct method for multidimensional ratio scaling.* Psychometrika, 28(1):33–41, 1963.

Chapter 4

(Un)suitable IFS Similarity Measures to Compare Experience-Based Evaluations

Abstract

In Chapter 3, we studied how to model a collection of experience-based evaluations (XBEs) by an intuitionistic fuzzy set (IFS). We also described how a similarity measure in the IFS framework can be augmented to obtain more reliable results when IFSs representing XBEs are compared to each other. As a sequel to that study, in this chapter we describe an empirical study whereby several similarity measures were tested in comparisons between pairs of IFSs that result from simulations of different evaluation processes. In such a simulation, the experience-based learning process described in Chapter 2 was used to learn how a human editor categorizes newswire stories under a specific scenario after which the resulting knowledge was used to evaluate the level to which other newswire stories fit into each of the learned categories. This chapter presents our findings about how (un)suitable each of the chosen similarity measures is for the comparison of the simulated XBEs.

This chapter is an adapted version of the following publications:

- Marcelo Loor and Guy De Tré. *Choosing Suitable Similarity Measures to Compare Intuitionistic Fuzzy Sets that Represent Experience-Based Evaluation Sets. Proceedings of the 7th International Joint Conference on Computational Intelligence*, 57-68. Lisbon, Portugal, 2015. This paper has been rewarded with the conference Best Student Paper Award.
- Marcelo Loor and Guy De Tré. *In a Quest for Suitable Similarity Measures to Compare Experience-Based Evaluations. Studies in Computational Intelligence*, edited by Juan Julian Merelo, Agostinho Rosa, José M. Cadenas, António Dourado, Kurosh Madani and Joaquim Filipe, 291-314 Vol. 669, Springer International Publishing, 2017.

4.1 Introduction

As explained in the previous chapter, a collection of XBEs can be characterized as an *intuitionistic fuzzy set* (IFS) [1, 2]. Thus, in theory, any of the existing similarity measures in the IFS framework can be used to compare such IFSs. However, we pointed out that, to compare two IFSs that represent XBEs, the similarity measures such as the studied in [3, 4] might not be applicable to this case because of their implicit assumption of symmetry, which does not adequately reflect judgments of similarity observed from a psychological perspective [5].

To determine which similarity measures in the IFS framework can be used for comparing XBEs, in this chapter we describe an empirical study in which such measures are tested by comparing pairs of IFSs characterizing simulated XBEs. Our motivation here is to complement the existing theoretical work within the context of IFSs to find suitable methods that allow us to compare collections of XBEs given by persons that might have different learning experiences.

To obtain simulated XBEs, we first use the experience-based learning method described in Section 2.2 to learn how a human editor categorizes newswire stories according to a given (learning) scenario. Then, we use the experience-based evaluation method described in Section 2.3 to evaluate the level to which newswire stories in a test collection fit into one of the learned categories.

Each of the established learning scenarios included a training collection that contains a certain proportion of opposite examples in relation to the original data, which consist of manually categorized newswire stories – by opposite example is meant that, e.g., if a story is assigned to a particular category in the original training collection, the story will not be assigned to the category in the training collection related to the current scenario.

An interesting aspect about testing the similarity measures in that way is that we can observe how each similarity measure reflects the perceived similarity between two collections of XBEs given from dissimilar learning scenarios. For instance, we were able to test a similarity measure to observe how it reflects the perceived similarity between the IFSs given by two persons who use training collections having examples that are totally opposite to each other – as was shown in Section 2.4, one can anticipate that the resulting level of similarity will be the lowest.

The remainder of this chapter is structured as follows. In the next section we present the similarity measures that were tested. After that, we describe how the collections of simulated XBEs were obtained in Section 4.3. Then, in Section 4.4 we describe the test procedure that was carried out for each of the chosen similarity measures. Before concluding this chapter, we present the results and our findings in Section 4.5.

4.2 Preliminaries

In the previous chapter we mentioned that, in the IFS framework a similarity measure, say S, is commonly used to compare IFSs. Thus, in theory, S can be used to compare two IFSs characterizing XBEs. For instance, consider that A and B are two IFSs that respectively represent the XBEs reflecting the experience or understanding that Alice and Bob have on what is a 'suitable comic book for 7-year-old kids'. Then, $S(A, B)$ will measure the similarity between those understandings.

In Section 3.4, the equations (3.14), (3.15), (3.16) and (3.17) were presented as examples of similarity measures based on a symmetric approach. Another such an example is a similarity measure based on Bhattacharyas's distance [3], which is defined by

$$S_{COS}(A,B)=\frac{1}{n}\sum_{i=1}^{n}\frac{\mu_A(x_i)\,\mu_B(x_i)+\nu_A(x_i)\,\nu_B(x_i)+h_A(x_i)\,h_B(x_i)}{\sqrt{\mu_A^2(x_i)+\nu_A^2(x_i)+h_A^2(x_i)}\sqrt{\mu_B^2(x_i)+\nu_B^2(x_i)+h_B^2(x_i)}}. \tag{4.1}$$

In addition, the following symmetric similarity measures that include the "notion of complement" in their definitions have been proposed in [6]:

$$S_{SK1}(A,B) = 1 - f\left(d(A,B), d(A,B^c)\right), \tag{4.2}$$

$$S_{SK2}(A,B) = \frac{1 - f\left(d(A,B), d(A,B^c)\right)}{1 + f\left(d(A,B), d(A,B^c)\right)}, \tag{4.3}$$

$$S_{SK3}(A,B) = \frac{\left(1 - f\left(d(A,B), d(A,B^c)\right)\right)^2}{\left(1 + f\left(d(A,B), d(A,B^c)\right)\right)^2} \tag{4.4}$$

and

$$S_{SK4}(A,B) = \frac{e^{-f(d(A,B),d(A,B^c))} - e^{-1}}{1 - e^{-1}}. \tag{4.5}$$

Herein, B^c is the complement of B, i.e.,

$$B^c = \{\langle x_i, \nu_B(x_i), \mu_B(x_i)\rangle | (x_i \in X) \wedge (0 \le \mu_B(x_i) + \nu_B(x_i) \le 1)\}, \tag{4.6}$$

$d(A, B)$ can be based on, e.g., the Hamming distance between A and B [4, 6], i.e.,

$$d(A,B)=\frac{1}{2n}\sum_{i=1}^{n}\left(|\mu_A(x_i)-\mu_B(x_i)|+|\nu_A(x_i)-\nu_B(x_i)|+|h_A(x_i)-h_B(x_i)|\right), \tag{4.7}$$

and

$$f\left(d(A,B), d(A,B^c)\right) = \frac{d(A,B)}{d(A,B) + d(A,B^c)}. \tag{4.8}$$

Equation (3.21) in Section 3.4 is another example of the formulation of a similarity measure using a directional approach. This similarity measure can be further augmented with a factor $\Delta_{@A} \in [0, 1]$ that indicates the level to which the understandings of the concept represented by IFSs A and B are in alignment. The augmented version can be defined by

$$S^{\alpha}_{@A}(A,B) = \Delta_{@A} \cdot S^{\alpha}(A,B). \tag{4.9}$$

As was discussed in Section 3.4, $\Delta_{@A}$ can be conceived as the weight of the *connotation-differential print* (CDP) between A and B as seen from the perspective of the evaluator who provides A.

4.3 Simulation

As previously stated, the aim of this chapter is to study empirically which of the similarity measures presented in the previous section can be used to compare collections of XBEs represented by IFSs. Hence, in this section we describe how the learning and the evaluation processes presented in Chapter 2 were used to obtain the IFSs that represent the collections of *simulated* XBEs.

4.3.1 Learning process

To simulate how a human editor categorizes newswire stories, we made use of the Reuters Corpora Volume I (RCV1) [7], which is a collection of manually categorized newswire stories provided by Reuters, Ltd. Specifically, we made use of the corrected version RCV1.v2, which is available (and fully described) in [8]. This collection has 804414 newswire stories, each assigned to one or more (sub) categories within three main categories: *Topics, Regions* and *Industries*.

The 23149 stories contained in the training file *lyrl2004_tokens_train.dat* of that collection were used to learn how to categorize newswire stories into one or more of the following 20 categories from *Topics*: *ECAT ("All Economics and Economic Indicators'), E11 ("Gross National/Domestic Product and economic performance"), E12 ("Monetary/Economic Policy and Intervention"), GSCI ("All aspects of science, research and new technology"), GSPO ("All sports stories"), GTOUR ("Tourism and tourist issues"), GVIO ("Civil unrest, demonstrations, civil war, war'), CCAT ("All Corporate-Industrial"), C12 ("Legal proceedings, court rulings and investigations"), C13 ("Regulation, deregulation, self regulation, rulings, government policy, licensing"), GCAT ("All Government and Social"), G15 ("All European Community affairs"), GDEF ("Defense policy"), GDIP ("Diplomatics affairs"), GDIS ("Natural Disasters"), GENT ("Entertainment"), GENV ("Environmental issues;"), GFAS ("Fashion related issues"), GHEA ("Health related issues")* and *GJOB ("Job related issues")* – the interested reader is referred to [8] for a full description of these categories.

To learn how to categorize newswire stories into each of the above categories, we established the following six learning scenarios:

- *R0*: All the stories in the training data are kept in their originally assigned training categories. So no modifications are introduced.

- *R20*, *R40*, *R60*, *R80*, *R100*: The original categorization is modified in, respectively, 20, 40, 60, 80 and 100% of the stories in the training data. The categorization in the remainder of the stories is preserved. The selection of the stories that obtain a modified categorization is made through a simple random sampling. Modification is done by reversing the membership of the story in its assigned categories. Hence, the opposite of its classification is obtained. For instance, consider a simulated learning process in which the scenario $R20$ is used for learning about the category *ECAT*. Consider also the story with code 2286, which is labeled as a member of the category *ECAT* in the original training data. If this story is one of the 20% of stories that have to modify its original categorization in this scenario, the story will be labeled as a nonmember of *ECAT*.

To use the learning method described in Section 2.2, each story has been represented by a vector whose elements are the words in the story. This follows an intuition in which, according to his/her experience, a person focuses on the words in a document to decide whether it fits or not into a given category – this intuition is similar to the one used in the *text categorization problem* [9], in which the words in a newswire story are the features that determine whether the story belongs or not to a category.

To simplify the vector representation, words like 'the', 'of' or 'at' that have a negligible impact on the categorization decision, and words like as 'learning', 'learned' or 'learn' that have a common stem can be filtered out and stemmed by using different algorithms. Hence, for the sake of reproducibility of the simulation, we made use of the stories in the training file *lyrl2004_tokens_train.dat* [8], which already have reduced and stemmed words. To obtain the reduced and stemmed words, in [8] techniques like tokenization, stop word removal, punctuation removal were applied to the text included in each newswire story. A token was defined as a "*maximal sequence of nonblank characters.*" Then, words included in the stop word list from the SMART system [10] were removed. After that, the tokens were stemmed by means of a particular implementation of the Porter stemmer [11]. For example, the story with code 2320 has the following words: *tuesday, stock, york, seat, seat, nys, level, million, million, million, sold, sold, current, off, exchang, exchang, exchang, bid, prev, sale, mln.* It is worth mentioning that the aim of a reduced collection of words is to decrease computational costs and increase categorization accuracy, which might be affected by redundant or irrelevant words.

To reflect the impact of the words on the categorization decision, the overall influence of each word should be assigned. Thus, as it was suggested in [8] we applied the equation

$$weight(f, x) = (1 + \ln n(f, x)) \ln \left(|X_0| / n(f, X_0) \right), \tag{4.10}$$

to compute the overall influence of a word in a story (or document). This equation is a variant of the *term frequency - inverse document frequency* weighting schema given in [12], where X_0 is the training collection (i.e., the collection of stories in *lyrl2004_tokens_train.dat*), $x \in X_0$ is a story, f is a word in x, $n(f, x)$ is the number of occurrences of f in x, $n(f, X_0)$ is the number of stories in X_0

that contain f, and $|X_0|$ is the number of stories in X_0 (i.e., $|X_0| = 23149$). For example, the overall influence of the word *exchang* in the story with code 2320 is given by $weight(exchang, 2320) = (1 + \ln 3) \ln (23149/2485) = 4.6834$.

After computing the overall influence of each word in a story, say x_i, we use the feature-influence representational model (see Section 2.2.2) to represent the resulting overall influence of the features of each $x_i \in X_0$ as a vector $\mathbf{x}_i = \beta_{i,1}\hat{\mathbf{f}}_1 + \cdots + \beta_{i,|\mathcal{F}|}\hat{\mathbf{f}}_{|\mathcal{F}|}$ such that:

- $\mathcal{F}$ is a dictionary containing all the distinct words in X_0;
- $|\mathcal{F}|$ is the number of words in $\mathcal{F}$ (for the chosen training collection, $|\mathcal{F}| = 47152$);
- $\hat{\mathbf{f}}_j$ is a unit vector that represents an axis related to a word $f_j \in \mathcal{F}$ (i.e., $\hat{\mathbf{f}}_j$ belongs to a multi-dimensional feature space in which each dimension corresponds to a word $f_j \in \mathcal{F}$); and
- $\beta_{i,j} = weight(f_j, x_i)$ is the weight of f_j in x_i (if f_j is not present in the story, $\beta_{i,j}$ will be fixed to 0). As was suggested in [8], each $\beta_{i,j}$ in $\mathbf{x}_i$ has been divided by $\|\mathbf{x}_i\| = \sqrt{\mathbf{x}_i \cdot \mathbf{x}_i}$.

Those vectors were used in the process described in Section 2.2.4 to compute a knowledge model $K_{C@LS} = \langle \hat{\mathbf{u}}_{C@LS}, t_{C@LS} \rangle$, for each category C under each established learning scenario LS as follows. First, we made use of the package *SVMLight Version V6.02* [13] to compute both (2.9) and (2.12) – in this case, we issued the command *"svm_learn.exe -c* 1 *svmTrainingFile svmModelFile,"* where *svmTrainingFile* is an input file that contains the training vectors for a category under a given scenario, and *svmModelFile* is an output file that contains the solution of the scenario-category learning process. After that, the results of (2.9) and (2.12) were used in (2.2) and (2.3) to compute $\hat{\mathbf{u}}_{C@LS}$ and $t_{C@LS}$ respectively.

Using the 6 learning scenarios and 20 categories described above, we obtained 120 *scenario-category* (knowledge) models – hereafter a knowledge model will be referred to using the nomenclature *scenario-category*.

4.3.2 Evaluation process

Consider a collection of newswire stories X. To evaluate the level to which a newswire story $x \in X$ fits into a category, say *ECAT*, under a given scenario, say *R20*, we use the *R20-ECAT* model, which represents the experience (or knowledge) acquired after following the learning process described in the previous section. After evaluating all the newswire stories in X, we obtain a simulated collection, say $ECAT_{@R20}$, that contains XBEs of which we consider that they are given by a person who learned the concept *ECAT* using the training data specified in the scenario *R20*.

The data and the process that were used to generate such collections of *simulated* XBEs are described below.

4.3.2.1 Evaluation data

We made use of the first 12500 newswire stories in each of the following files from RCV1.v2 [8]:

- *lyrl2004_tokens_test_pt0.dat*,
- *lyrl2004_tokens_test_pt1.dat*,
- *lyrl2004_tokens_test_pt2.dat* and
- *lyrl2004_tokens_test_pt3.dat*.

With these 50000 stories, we built 1000 50-story collections.

4.3.2.2 Obtaining an IFS as a result of an evaluation process

Let X_k be one of the 50-story collections that constitute the evaluation data. To evaluate the level to which a story $x_i \in X_k$ fits into a category, say C, under a given (learning) scenario, say LS, we made use of the knowledge model $K_{C@LS}$ resulting from the previous learning process to obtain an IFS element $\langle x_i, \mu_C(x_i), \nu_C(x_i)\rangle$ as follows.

First, we represented $x_i \in X_k$ by a vector $\mathbf{x}_i = \beta_{i,1}\hat{\mathbf{f}}_1 + \cdots + \beta_{i,|\mathcal{F}|}\hat{\mathbf{f}}_{|\mathcal{F}|}$ according to the procedure described in the previous section, where X_0 corresponds to the training collection in the scenario LS.

Then, we made use of $\hat{\mathbf{u}}_{C@LS} = \omega_1\hat{\mathbf{f}}_1 + \cdots + \omega_{|\mathcal{F}|}\hat{\mathbf{f}}_{|\mathcal{F}|}$ and $t_{C@LS}$ in $K_{C@LS}$ to compute $\mu_{C@LS}(x_i)$ and $\nu_{C@LS}(x_i)$ by means of the equations

$$\mu_{C@LS}(x_i) = \breve{\mu}_{C@LS}(x_i)/\eta \tag{4.11}$$

and

$$\nu_{C@LS}(x_i) = \breve{\nu}_{C@LS}(x_i)/\eta \tag{4.12}$$

respectively, where

$$\breve{\mu}_{C@LS}(x_i) = \begin{cases} \frac{\left(\sum_{j=1}^{|\mathcal{F}|} \beta_{i,j}\omega_j\right) + |t_{C@LS}|}{\|\mathbf{x}_i\|} & \text{if } (\forall j : \beta_{i,j}\omega_j > 0) \wedge (t_{C@LS} < 0)\,; \\ \frac{\sum_{j=1}^{|\mathcal{F}|} \beta_{i,j}\omega_j}{\|\mathbf{x}_i\|} & \text{if } (\forall j : \beta_{i,j}\omega_j > 0) \wedge (t_{C@LS} \geq 0)\,; \\ 0 & \text{otherwise;} \end{cases} \tag{4.13}$$

$$\breve{\nu}_{C@LS}(x_i) = \begin{cases} \frac{\left(\sum_{j=1}^{|\mathcal{F}|} |\beta_{i,j}\omega_j|\right) + t_{C@LS}}{\|\mathbf{x}_i\|} & \text{if } (\forall j : \beta_{i,j}\omega_j < 0) \wedge (t_{C@LS} > 0) \\ \frac{\sum_{j=1}^{|\mathcal{F}|} |\beta_{i,j}\omega_j|}{\|\mathbf{x}_i\|} & \text{if } (\forall j : \beta_{i,j}\omega_j < 0) \wedge (t_{C@LS} \leq 0)\,; \\ 0 & \text{otherwise;} \end{cases} \tag{4.14}$$

and

$$\eta = \max\left(1, \breve{\mu}_{C@LS}(x_i) + \breve{\nu}_{C@LS}(x_i)\right), \forall x_i \in X_k. \tag{4.15}$$

Finally, after computing the IFS elements for each $x_i \in X_k$, we obtained an IFS that represents the simulated XBEs for the stories in X_k according to what was learned (or experienced) about the category C under the scenario LS.

Table 4.1: IFSs that represent the simulated experience-based evaluations for the stories in each $X_k \in \{X_1, \cdots, X_{1000}\}$ according to what was learned about category *E11* under the scenarios *R0, R20, R40, R60* and *R100* respectively.

	E11				
	50-story Collections				
Scenario	X_1	$\cdots$	X_k	$\cdots$	X_{1000}
R0	$E11_{@R0}(X_1)$	$\cdots$	$E11_{@R0}(X_k)$	$\cdots$	$E11_{@R0}(X_{1000})$
R20	$E11_{@R20}(X_1)$	$\cdots$	$E11_{@R20}(X_k)$	$\cdots$	$E11_{@R20}(X_{1000})$
R40	$E11_{@R40}(X_1)$	$\cdots$	$E11_{@R40}(X_k)$	$\cdots$	$E11_{@R40}(X_{1000})$
R60	$E11_{@R60}(X_1)$	$\cdots$	$E11_{@R60}(X_k)$	$\cdots$	$E11_{@R60}(X_{1000})$
R80	$E11_{@R80}(X_1)$	$\cdots$	$E11_{@R80}(X_k)$	$\cdots$	$E11_{@R80}(X_{1000})$
R100	$E11_{@R100}(X_1)$	$\cdots$	$E11_{@R100}(X_k)$	$\cdots$	$E11_{@R100}(X_{1000})$

Since we built 1000 50-story collections, we obtained 1000 IFSs for each scenario-category model. We made use of the notation $C_{@LS}(X_k)$ to denote an IFS that represents the simulated XBEs for the stories in X_k according to what was learned about category C under a scenario *LS*. For example, Table 4.1 shows an extract of the IFSs that represent the simulated XBE for the stories in each $X_k \in \{X_1, \cdots, X_{1000}\}$ according to what was learned about category *E11* under the scenarios *R0, R20, R40, R60* and *R100* respectively.

Considering that we chose 20 categories and built 6 scenarios during the learning phase, we obtained a total of 120000 IFSs during this phase.

4.4 Testing

In this section we describe how the similarity measures presented in Section 4.2 were tested with the IFSs that represent the collections of simulated XBEs.

4.4.1 A point of reference for the perceived similarity

Consider a scenario-category model LS-C represented by both $\mathbf{w}$ and b according to the equations (2.9) and (2.12) respectively. Consider then a story $x_i \in X_k$ represented by $\mathbf{x_i}$, where X_k is one of the 50-story collections in the evaluation data. Consider finally a collection $Y_k = \{y_i | (y_i = \mathbf{w} \cdot \mathbf{x_i} + b)\}$ such that y_i is the *SVM-based evaluation* of story $x_i \in X_k$ fitting into the category C under the scenario LS. In this context, the decision about the fittingness of the story x_i into the category C under the scenario LS will depend on y_i: when $y_i > 0$, the decision will be "x_i fits into C;" when $y_i < 0$, the decision will be "x_i does not fit into C;" and when $y_i = 0$, no decision will be taken.

Now consider the collections $Y_{k@L1}$ and $Y_{k@L2}$ having SVM-based evaluations under scenarios $L1$ and $L2$ respectively. Consider also $y_{i@L1} \in Y_{k@L1}$ and $y_{i@L2} \in Y_{k@L2}$. In this situation, when

$$((y_{i@L1} < 0 \wedge y_{i@L2} < 0) \vee (y_{i@L1} > 0 \wedge y_{i@L2} > 0) \vee (y_{i@L1} = 0 \wedge y_{i@L2} = 0))$$

is *true*, an *agreement on decision* about the fittingness of story x_i between the evaluations given under scenarios $L1$ and $L2$ occurs.

We made use of the agreements on decisions between $Y_{k@L1}$ and $Y_{k@L2}$ to obtain an *agreement-on-decision ratio*, AoD for short, which is expressed by

$$AoD(Y_{k@L1}, Y_{k@L2}) = n/N, \tag{4.16}$$

where n represents the number of agreements on decision between $Y_{k@L1}$ and $Y_{k@L2}$, and N represents the number of stories in X_k. Since the AoD ratio denotes how similar the decisions are, we deemed it to be an indicator of the perceived similarity between the evaluations given by two persons that learned (or experienced) C under $L1$ and $L2$ respectively.

4.4.2 Testing procedure and settings

As has been mentioned throughout this dissertation, an XBE mainly depends on what an evaluator has experienced or learned about a particular concept. Thus, one could expect that the level of similarity between the collections of XBEs given by two evaluators who learned a concept under the same (learning) scenario will be greater than or equal to the level of similarity between the collections of XBEs given by two evaluators who learned the same concept under different scenarios. For instance, consider three evaluators: P, Q and R. While P and Q learned about the category $E11$ under the same scenario $R0$, R learned so under the scenario $R80$. Consider also that the IFSs $E11_{@P}(X_k)$, $E11_{@Q}(X_k)$ and $E11_{@R}(X_k)$ represent the collections of XBEs about the fittingness of the stories in the 50-story collection X_k into category $E11$ given by P, Q and R respectively. In this context, one could expect that the similarity between $E11_{@P}(X_k)$ and $E11_{@Q}(X_k)$ will be greater than the similarity between $E11_{@P}(X_k)$ and $E11_{@R}(X_k)$.

We made use of the above intuition to test the similarity measures presented in Section 4.4.2. Since we chose the AoD ratio as an indicator of the perceived similarity, we first tested it to observe how the agreement on decisions between two SVM-based evaluation sets is affected according to their respective learning scenarios. We then tested the similarity measures, some of them with different configurations as will be described next.

4.4.2.1 Testing the agreement-on-decision ratio

Under the same assumptions, one could expect that the AoD ratio between two SVM-based evaluation sets resulting from the same scenario will be greater than the AoD ratio between two SVM-based evaluation sets resulting from different scenarios. Thus, we considered the question: *Is there sufficient evidence in the evaluation data to suggest that the mean AoD ratio changes after altering a given percentage of the training data?* To answer this, for each category and for each 50-story collection, we obtained the AoD ratio between the SVM-based evaluation set given under scenario $R0$ (i.e., $R0$ is a *referent* scenario) and each of the SVM-based evaluation sets given under the scenarios $R0, R20, R40, R60, R80$ and $R100$ respectively (see Algorithm 1).

Algorithm 1: Obtaining AoD ratios

Require: $ChosenCategories$ // see Section 4.3.1
Require: $LearningScenarios$ // see Section 4.3.1
Require: $50storyCollections$ // see Section 4.3.2.1
Require: $SVMEvals$ // see Section 4.4.1
$Z \leftarrow \emptyset$ // resulting ratios
for all $C \in ChosenCategories$ **do**
  for all $X_k \in 50storyCollections$ **do**
    $Y_{k@R0} \leftarrow SVMEvals[X_k][R0][C]$
    for all $LS \in LearningScenarios$ **do**
      $Y_{k@LS} \leftarrow SVMEvals[X_k][LS][C]$
      $r \leftarrow AoD(Y_{k@R0}, Y_{k@LS})$
      $Z[C][LS][X_k] \leftarrow r$
return Z

4.4.2.2 Testing the similarity measures

To test the similarity measures, we computed the level of similarity between the IFS given under scenario $R0$ and each of the IFSs given under the scenarios $R0, R20, R40, R60, R80$ and $R100$ respectively by means of each of the established similarity measures. We did so through the steps described in Algorithm 2. As could be noticed, the computation was performed for each category, for each 50-story collection, for each scenario and for each similarity measure. For readability, hereafter we shall use the acronym placed as subscript in each of the given similarity measures to refer to each of them. For instance, we shall use *H2D* to refer to S_{H2D} (see (3.15)).

Since two of the similarity measures presented in Section 4.2, namely S^{α} (see (3.21)) and its augmented version (see (4.9)), needed to be configured, we used the configurations described next to perform the test.

The similarity measure S^{α} was configured with hesitation splitters $\alpha = 0$, 0.5 and 1 – we shall use the label *VB-*α to refer to each of the used configurations for this measure.

With respect to the extended version of S^{α}, two different methods were applied to compute the $\Delta_{@A}$ factor, i.e., two forms of this measure were used during the test.

In the first form, labeled *XVB-*α*-*w, $\Delta_{@A}$ was computed by means of the method $weightCDP(A, B, \alpha, w)$, in which A and B are the IFSs in the comparison, α is the hesitation splitter, and $w \in [0, 1]$ is a value that allows us to obtain a CDP (see Section 4.3.2.2) between A and B according to the *width* of the average gap between the membership and non-membership values as seen from the perspective of the provider of A. This method involves the following steps:

Algorithm 2: Testing similarity measures

Require: $SimMeasures$ // see Section 4.4.2.2
Require: $ChosenCategories$ // see Section 4.3.1
Require: $LearningScenarios$ // see Section 4.3.1
Require: $50storyCollections$ // see Section 4.3.2.1
Require: $IFSEvals$ // see Section 4.3.2.2
$Z \leftarrow \emptyset$ // resulting levels
for all $C \in ChosenCategories$ **do**
  for all $X_k \in 50storyCollections$ **do**
    $C_{@R0}(X_k) \leftarrow IFSEvals[X_k][R0][C]$
    for all $LS \in LearningScenarios$ **do**
      $C_{@LS}(X_k) \leftarrow IFSEvals[X_k][LS][C]$
      for all $S \in SimMeasures$ **do**
        $l \leftarrow S(C_{@R0}(X_k), C_{@LS}(X_k))$
        $Z[C][LS][X_k][S] \leftarrow l$
return Z

1. Obtain $\delta \in [0, 1]$ for IFS A through

$$\delta = \frac{w}{n} \sum_{i=1}^{n} (\mu_A(x_i) + \nu_A(x_i)). \tag{4.17}$$

2. Compute the spot differences among the IFS elements in A and B using (3.10).
3. Order the IFS elements in A by descending membership values. If two IFS elements have the same membership value, order them by their ascending non-membership values.
4. Initialize $k = \frac{n}{10}$ (i.e., $k = 5$) and obtain the connotation-differential markers (i.e., ⟠, ⍏ and ⍖) for the k-highest and the k-lowest IFS elements in the ordered IFS A (see Figure 4.1). For a spot difference s, the marker will be: ⟠ when $|s| \leq \delta$; ⍏ when $s > \delta$; and ⍖ when $s < -\delta$.
5. Build the CDPs cdp_H and cdp_L with the markers corresponding to k-highest and the k-lowest IFS elements respectively (see Figure 4.1).
6. Set $v[$⟠$] = 1$, $v[$⍏$] = 0.01$ and $v[$⍖$] = 0.01$, and compute $\Delta_{@A}$ by means of

$$\Delta_{@A} = \max \left(\frac{1}{k} \sum_{m \in cdp_H} v[m], \frac{1}{k} \sum_{m \in cdp_L} v[m] \right) \tag{4.18}$$

In the second form, labeled *XVBr-α*, $\Delta_{@A}$ was computed through a novel method, called *spotRatios*, which involves the following steps:

1. Initialize $k = \frac{n}{10}$ (i.e., $k = 5$).

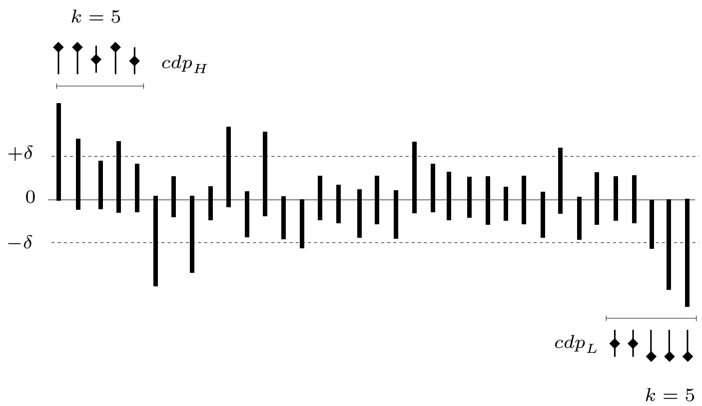


Figure 4.1: Obtaining a CDP and its weight. The bars represent the spot differences between the elements of IFSs A and B. The CDPs for the k-highest and the k-lowest IFS elements according to A's perspective are denoted by cdp_H and cdp_L respectively.

2. Order the IFS elements in A by descending membership values. If two IFS elements have the same membership value, order them by their ascending non-membership values. After that, put the top k into a collection H and the bottom k in a collection L.

3. Compute $\Delta_{@A}$ by means of

$$\Delta_{@A} = \frac{1}{2k}\left(\sum_{\mathbf{a}_i \in H} spotRatio(\mathbf{a}_i, \mathbf{b}_i, \alpha) + \sum_{\mathbf{a}_i \in L} spotRatio(\mathbf{a}_i, \mathbf{b}_i, \alpha)\right), \tag{4.19}$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are the vector representations of the IFS elements in A and B related to x_i (see Section 3.3.1), α is the hesitation splitter, and *spotRatio* is defined by Algorithm 3. In this algorithm, we obtain first $\mathbf{o}_i$, which is the vector representation of the complement of the IFS element represented by $\mathbf{a}_i$ (see Lines 6 and 7). Then, we compute A_{ao}, which is the area of the parallelogram formed by $\mathbf{a}_i$ and $\mathbf{o}_i$ (see Line 8), as well as A_{bo}, which is the area of the parallelogram formed by $\mathbf{b}_i$ and $\mathbf{o}_i$ (see Line 9) – these areas are depicted in Figure 4.2. After that, the ratio r between A_{bo} and A_{ao} is computed (see Line 11). This ratio is returned after validating the upper and the lower bounds in Lines 13 and 15 respectively.

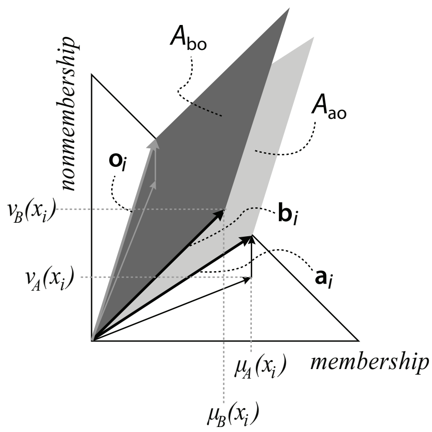


Figure 4.2: Computing a spot ratio.

Algorithm 3: *spotRatio*$(\mathbf{a}_i, \mathbf{b}_i, \alpha)$

$r \leftarrow 0.5$ // default value
$a_\mu \leftarrow \mu_A(x_i) + \alpha h_A(x_i)$
$a_\nu \leftarrow \nu_A(x_i) + (1-\alpha) h_A(x_i)$
$b_\mu \leftarrow \mu_B(x_i) + \alpha h_B(x_i)$
$b_\nu \leftarrow \nu_B(x_i) + (1-\alpha) h_B(x_i)$
$o_\mu \leftarrow \nu_A(x_i) + \alpha h_A(x_i)$ // $\mathbf{o}_i$ is the vector representation of the complement of the IFS element $\langle x_i, \mu_A(x_i), \nu_A(x_i) \rangle$ represented by $\mathbf{a}_i$
$o_\nu \leftarrow \mu_A(x_i) + (1-\alpha) h_A(x_i)$
$A_{ao} \leftarrow a_\mu o_\nu - a_\nu o_\mu$ // A_{ao} is the area of the paralellogram formed by $\mathbf{a}_i$ and $\mathbf{o}_i$
$A_{bo} \leftarrow b_\mu o_\nu - b_\nu o_\mu$ // A_{bo} is the area of the paralellogram formed by $\mathbf{b}_i$ and $\mathbf{o}_i$
if $|A_{ao}| > 0$
    $r \leftarrow A_{bo}/A_{ao}$
    if $r > 0$
        $r \leftarrow \min(1, r)$
    else
        $r \leftarrow 0$
return r

4.5 Results and Practical Implications

This section presents the experimental results that were obtained using the test conditions described in the previous section. Also, it presents some practical implications of these results.

4.5.1 Agreement-on-decision ratio as an indicator of the perceived similarity

To answer the question "*is there sufficient evidence in the evaluation data to suggest that the mean AoD ratio changes after altering a given percentage of the training data?*," we first made use of the collection resulting of Algorithm 1 to compute the averages of the AoD ratios per scenario-category. We then ran the t-test for the null hypothesis "*the average of the AoD ratio is the same after altering the r% of the training data*" in contrast to the alternative one "*the average of the AoD ratio changes after altering the r% of the training data*" according to r given in each scenario (see Table 4.2).

Table 4.2: Averages of the AoD ratios per scenario-category, and t-test for the null hypothesis "the average of the AoD ratio is the same after altering the r% of the training data" according to r given in each scenario (e.g., $r = 20$ in scenario *R20*), where *R0* ($r = 0$) is the referent scenario.

	Scenarios				
Category	R20	R40	R60	R80	R100
C12	0.7292	0.5900	0.4757	0.2897	0.0001
C13	0.7385	0.6091	0.4281	0.2766	0.0002
CCAT	0.9372	0.7505	0.2711	0.0663	0.0001
E11	0.6431	0.5740	0.4722	0.3519	0.0003
E12	0.7156	0.5792	0.4796	0.3080	0.0001
ECAT	0.8187	0.6273	0.4052	0.1853	0.0002
G15	0.7186	0.5954	0.4781	0.3039	0.0002
GCAT	0.9314	0.7472	0.2690	0.0661	0
GDEF	0.6515	0.5717	0.4668	0.3602	0.0002
GDIP	0.7433	0.5990	0.4587	0.2672	0.0002
GDIS	0.7229	0.5951	0.4729	0.3022	0.0002
GENT	0.7066	0.5796	0.4802	0.3209	0.0002
GENV	0.6941	0.6009	0.4812	0.3248	0.0004
GFAS	0.6763	0.5787	0.4882	0.3457	0.0002
GHEA	0.7016	0.5850	0.4636	0.3451	0.0003
GJOB	0.7359	0.5883	0.4542	0.2886	0.0003
GSCI	0.6899	0.5854	0.4844	0.3424	0.0004
GSPO	0.8208	0.6508	0.4130	0.1962	0
GTOUR	0.5383	0.5197	0.4866	0.4663	0.0005
GVIO	0.7551	0.6368	0.4749	0.2796	0.0002
Mean	0.7334	0.6082	0.4452	0.2844	0.0002
stdDev	0.0913	0.0551	0.0643	0.0951	0.0001
N	20	20	20	20	20
df	19	19	19	19	19
t-value	13.06	31.80	38.59	33.67	34025.85
p-value	0	0	0	0	0

The results in Table 4.2 show that, for the scenarios *R20, R40, R60, R80* and *R100*, the t-values were statistically significant ($p < 0.05$). Consequently, we can say that there is sufficient evidence in the evaluation data to suggest that the average of the AoD ratio changes after altering the 20, 40, 60, 80 or 100% of the training data.

Recalling that we deemed the AoD ratio to be an indicator of the perceived similarity, we can confidently expect that it will be affected by the different learning scenarios established in the simulation. This can be observed in the bivariate plot depicted in Figure 4.3, which shows a strongly negative (or inverse) relationship ($R = -0.9741$) between the averages of the AoD ratios and the percentage of opposites included in the learning scenarios.

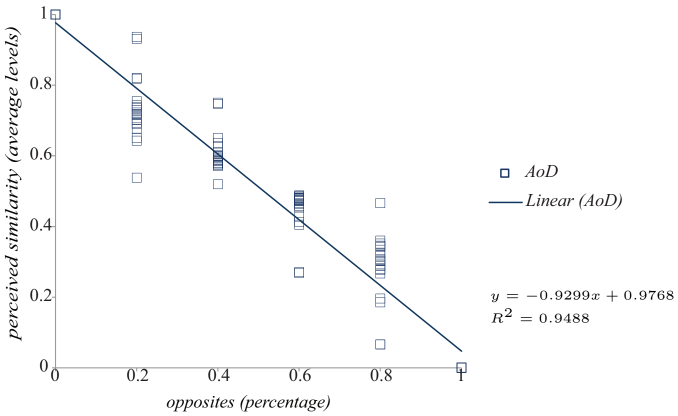


Figure 4.3: Bivariate plot between the averages of the AoD ratios and the percentage of opposites included in the learning scenarios. The relationship is represented by means of a linear model and described by the statistic R (Pearson Product Moment Correlation).

4.5.2 To what extent does each similarity measure reflect the perceived similarity?

To check to what extent each of the configurations of similarity measures given in Section 4.4.2.2 reflects the perceived similarity between the simulated IFSs, we first made use of the collection resulting of Algorithm 2 to compute the averages of the levels of similarity per scenario-category. Then, we obtained linear models for the relationships between each one of those averages and the percentage of opposites considered in each scenario. After that, each of the resulting models was contrasted with the linear model corresponding to the AoD ratio. As an indicator of how well a similarity measure reflects the perceived similarity, we computed a *manifest* index, which is defined by

$$m = (a_{SM}/a_{AoD})(b_{SM}/b_{AoD})(R^2_{SM}/R^2_{AoD}), \tag{4.20}$$

where a_{SM} and a_{AoD} are the slopes, b_{SM} and b_{AoD} are the intercepts, and R^2_{SM} and R^2_{AoD} are the R-statistics in the linear models corresponding to the similarity measure SM and the AoD ratio respectively. For readability, we shall use hereafter *SM-vs.-OP* to denote the relationship between the averages of the levels (of similarity) resulting from the (configuration of) similarity measure SM and the percentage of opposites OP.

Table 4.3: Linear models and m-indices for each *SM-vs.-OP* representing the relationship between the averages levels that result from the (configuration of) similarity measure SM and the percentage of opposites OP.

	SM-vs.-OP (linear model: $y = ax + b$)			
SM	slope (a)	intercept (b)	R^2	m-index
H2D	−0.0139	0.9939	0.4128	0.0066
H3D	−0.0138	0.9852	0.1442	0.0023
E2D	−0.0171	0.9920	0.4189	0.0082
E3D	−0.0167	0.9853	0.2034	0.0039
COS	−0.0004	0.9831	0	0
SK1	**−0.7302**	**0.8666**	**0.7451**	**0.5471**
SK2	−0.7287	0.7547	0.6920	0.4416
SK3	−0.7242	0.6041	0.5242	0.2661
SK4	−0.7298	0.7848	0.7163	0.4761
VB-0	−0.0133	0.9955	0.4527	0.0070
VB-0.5	−0.0133	0.9955	0.4528	0.0070
VB-1	−0.0144	0.9922	0.3170	0.0053
XVB-0-0.05	−0.7318	0.7831	0.6871	0.4569
XVB-0.5-0.05	−0.6738	0.6999	0.5974	0.3269
XVB-1-0.05	−0.6388	0.6185	0.4666	0.2139
XVB-0-0.1	−0.6587	0.9307	0.6560	0.4666
XVB-0.5-0.1	−0.6240	0.8358	0.6878	0.4162
XVB-1-0.1	−0.5727	0.6978	0.4805	0.2228
XVB-0-0.2	−0.4218	1.0241	0.4575	0.2293
XVB-0.5-0.2	−0.4657	1.0029	0.5944	0.3221
XVB-1-0.2	−0.4335	0.8321	0.4414	0.1847
XVBr-0	−0.7368	0.8069	0.6740	0.4650
XVBr-0.5	**−0.7971**	**0.8984**	**0.8499**	**0.7062**
XVBr-1	−0.7368	0.8069	0.6740	0.4650
AoD	**−0.9299**	**0.9768**	**0.9488**	**1**

The results in Table 4.3 show that, in contrast to what happens with the AoD ratio, the averages of the levels of *H2D*, *H3D*, *E2D*, *E3D*, *COS*, *VB-0*, *VB-0.5* and *VB-1* are hardly affected by the variation of the percentage of opposites – notice the broad difference among the slopes of the linear models corresponding to these similarity measures and the slope of the linear model corresponding to the AoD ratio. By way of illustration, according to the results in Table 4.3 the slope and the intercept term for *COS* are −0.0004 and 0.9831 respectively. Hence, the linear model computed for *COS* is $y = -0.0004x + 0.9831$ (see Figure 4.4). If we use this linear model to compute the level to which the average of evaluations given under the scenarios $R0$ and $R100$ are similar, i.e., $y = -0.0004(1) + 0.9831$, we will obtain $y = 0.9827$ as a result even though a result close to 0 is expected in this case – since $R100$ contains 100% of opposite training examples in relation to $R0$, we set $x = 1$ to make this computation. This result shows that *COS* is hardly affected by the variation of the percentage of opposites. In other words, the previous result shows that *COS* does not properly reflect the perceived similarity when XBEs resulting from totally opposite learning scenarios are compared to each other.

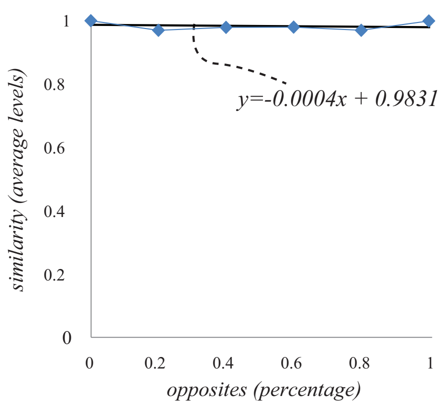
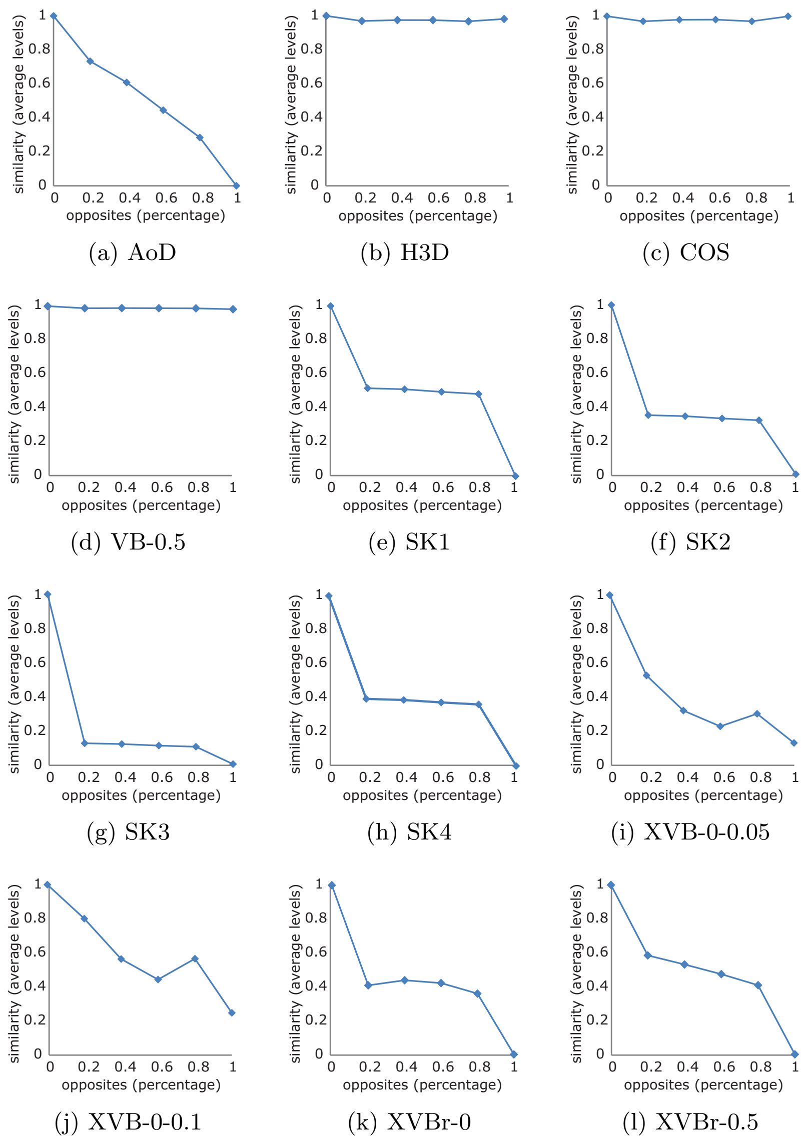


Figure 4.4: Linear model computed for the *COS* similarity measure.

Regarding the averages of the levels of *SK1*, *SK2*, *SK3* and *SK4*, the results suggest that the levels computed with *SK1*, *SK2* and *SK4* are fairly affected by the variation of the percentage of opposites. Notice that the correlation indices for *SK1-vs.-OP*, *SK2-vs.-OP* and *SK4-vs.-OP* (i.e., $R = -0.8632$, $R = -0.8319$ and $R = -0.8463$ respectively) denote fairly strong negative relationships that are roughly comparable with the strongly negative relationship ($R = -0.9741$) in *AoD-vs.-OP*. Moreover, Figures 4.5e, 4.5f and 4.5h show that *SK1*, *SK2* and *SK4* reflect properly the perceived similarity between the evaluations given under the scenarios $R0$ and $R100$ in contrast to, e.g., *H3D*, *COS* or *VB-0.5* (see Figures 4.5b, 4.5c and 4.5d). However, these similarity measures seem to reflect more or less properly the perceived similarity between the evaluations given under the scenarios $R0$ and, e.g., $R20$ or $R80$, which affects the values of the m-indices related to their linear models.

With respect to the averages of the levels of the form *XVB*-α-w of (4.9), the results in Table 4.3 show that the levels computed with two of them, namely *XVB-0-0.05* and *XVB-0-0.1*, are fairly affected by the variation of the percentage of opposites as well. In contrast to *SK1*, *SK2* and *SK4*, Figures 4.5i and 4.5j suggest that *XVB-0-0.05* and *XVB-0-0.1* reflect more properly the perceived similarity between the evaluations given under the scenario $R0$ and the evaluations given under the scenarios $R20$, $R40$ or $R60$. However, the figures



(a) AoD (b) H3D (c) COS

(d) VB-0.5 (e) SK1 (f) SK2

(g) SK3 (h) SK4 (i) XVB-0-0.05

(j) XVB-0-0.1 (k) XVBr-0 (l) XVBr-0.5

Figure 4.5: Averages of the similarity levels per scenario versus the percentage of opposites included in each scenario.

also suggest that both measures do not properly reflect the perceived similarity between the evaluations given under the scenarios $R0$ and $R80$ or $R100$ – notice that the average of the similarity levels between $R0$ and $R80$ is greater that the average of the similarity levels between $R0$ and $R60$.

Since the form XVB-α-w is based on the weight of a CDP and the computa-

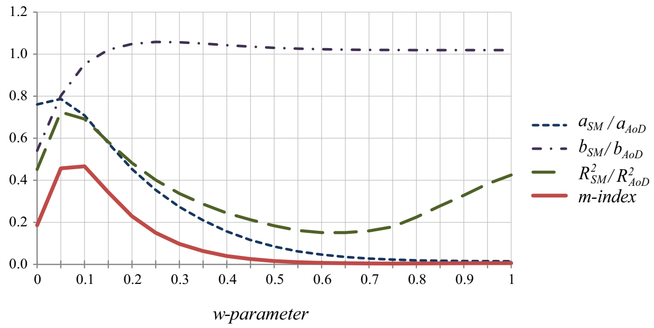


Figure 4.6: Influence of the w-parameter on the quality of the m-index for XVB-0-w.

tion of this weight has been based on the w-parameter in our testing procedure (see Section 4.4.2.2), we performed additional tests to observe the influence of this parameter on the quality of the results of this similarity measure. In such additional tests, we configured (4.9) with $\alpha = 0$ and $w = 0.05, 0.1, 0.15, \cdots, 1$ and used the same nomenclature (i.e., *XVB-α-w*) to label each configuration. Figure 4.6 shows how the m-index corresponding to the linear model for each *XVB-0-w-vs-OP* relationship is affected by the w-parameter. As noticed, the peak m-index is reached at $w = 0.1$ and is projected to decline after that point. As was mentioned in Section 4.4.2.2, the w-parameter determines the *width* of the average gap between the membership and non-membership values, which is then used to build a CDP for the IFSs in the similarity comparison as seen from the perspective of the person who provides the referent IFS. This means that, in this scenario, a spot difference with a magnitude less than or equal to the 10% of the average gap between the membership and non-membership values will roughly reflect a similar understanding (or knowledge) of the evaluate concept. This result seems to support the idea behind a CDP, which suggests that a difference in understanding of a concept could be marked by a difference in one or more evaluations of relevant objects.

Concerning the averages of the levels of the form $XVBr$-α of (4.9), the results reported in Table 4.3 show that the levels computed by $XVBr$-0.5 are strongly affected by the variation of the percentage of opposites. As noticed, the correlation index for $XVBr$-0.5-*vs.-OP* ($R = -0.9219$) denote a very strong inverse relationship that is comparable with the correlation index for *AoD-vs.-OP* ($R = -0.9741$). In addition, Figure 4.5l shows that the averages of the similarity levels between the evaluations given under scenario *R0* and the ones given under the other scenarios are well reflected by $XVBr$-0.5, which is indicated by the best m-index reported (i.e., $m = 0.7062$). However, with respect to the levels computed by $XVBr$-0 and $XVBr$-1, the results suggest that such levels are affected by the variation of the percentage of opposites but not as strong as width $XVBr$-0.5. A potential weakness of $XVBr$-0 when compared to $XVBr$-0.5 is shown in Figure 4.5k. Notice that, in contrast to $XVBr$-0.5, the average of the similarity levels computed by $XVBr$-0 between $R0$ and $R20$ is a little less than the average computed between $R0$ and $R40$.

4.5.3 Discussion

The results suggest that one of the configurations of the similarity measure (4.9), namely *XVBr-0.5*, outperforms the other (configurations of) similarity measures when dealing with similarity comparisons among the simulated IFSs. However, it was found that other (configurations of) similarity measures such as *XVB-0-0.1* or *SK1* can be (partially) effective in comparisons between IFSs resulting from particular scenarios. For instance, *SK1* seems to reflect properly the perceived similarity between the evaluations resulting from completely opposite understandings but it reflects to a lesser extent the perceived similarity among the evaluations resulting from roughly opposite (or slightly similar) understandings.

A possible explanation for those results might be that, by means of the factor $\Delta_{@A}$, the configurations of the similarity measure (4.9) take into account what is understood as a qualitative difference between two IFS elements from the perspective of the evaluator who provides the IFS A. This situation is observable in the two evaluated forms of this measure: in the form *XVB-*α*-*w, when both the magnitude and the direction of a spot difference (see Section 3.3.2), as well as the average gap between the membership and non-membership components of the IFS elements in A are used in the computation of $\Delta_{@A}$ (see (4.18)); and in the form *XVBr-*α, when both the magnitude and the direction of a vector product $\mathbf{a}_i \times \mathbf{o}_i$, in which $\mathbf{a}_i$ is a vector representing an IFS element in A and $\mathbf{o}_i$ is a vector representing the complement of that IFS element, are used as points of reference in the (internal) computation of $\Delta_{@A}$ (see Algorithm 3). Even though both forms try to detect and quantify any qualitative difference between two IFS elements, the results suggest that the form *XVBr-*α is a more effective. Notice that, in contrast to the form *XVB-*α*-*w, the form *XVBr-*α does not need a threshold value to quantify a difference (or similarity) in understandings, i.e., the parameter w is not necessary. Notice also that the method applied by *XVBr-*α seems to agree in some extent with the "notion of complement" used in the definitions of *SK1*, *SK2*, *SK3* and *SK4*.

Another possible explanation for the results might be that a gap between the membership and non-membership components is contextually related to the categorization decision (see Sections 4.3.2.2 and 4.4.1), which is deemed to be a point of reference for the perceived similarity through the agreement on the decision ratio. Hence, a similarity measure such as (4.9) that takes into account the aforesaid gap could reflect more adequately the similarity perceived from the perspective of who makes the categorization decision.

Even though these results are based on simulated IFSs that use a manually categorized newswire stories, they need to be interpreted with caution because of the dependency of the IFSs with the learning algorithm and the (text categorization) context that were chosen for the simulations. Consequently, conducting simulations with other learning algorithms and experiments with real evaluators is recommended and subject to further study.

4.5.4 Practical implications

IFSs have been applied to solve problems in topics like decision making, see e.g., [6, 14, 15] or pattern recognition, see e.g., [16, 17, 18]. Since similarity measures aiming to compare IFSs are required to solve those problems, the results reported above have practical implications for the solution of them.

An important implication is that processes like querying, decision making or data mining in which such similarity measures are used, can be improved. In this way, business activities that use or depend on the aforementioned processes (e.g., marketing analysis, identification of patterns or consensus reaching on the development of new products or services) will benefit from the reported results because the users are provided with better information on context. For instance, we foresee that the identification of patterns on what customers experience or perceive about products or services can be improved by using the kind of contextualized comparison results obtained with the techniques developed in this chapter. Same observations hold for all areas where IFSs have been applied.

4.6 Conclusions

In this chapter we have described an empirical study that aims to determine which similarity measures are suitable to compare experience-based evaluations (XBEs) represented by intuitionistic fuzzy sets (IFSs).

During the study, several similarity measures proposed to compare IFSs were used in comparisons of simulated XBEs, which were obtained after learning through a support vector learning algorithm how human editors categorize newswire stories under different scenarios. This made it possible to assess the level to which each similarity measure reflects the perceived similarity in comparisons of XBEs that might be given by persons with different backgrounds. In this regard, this study can be seen as an integrated package that helps to answer Research Questions $Q2$, $Q3$, $Q4$ and $Q5$.

Taken together, the results presented in this chapter suggest that the studied similarity measures could be categorized as *suitable*, *partially suitable* and *unsuitable* while comparing XBEs.

The first category includes an improved version of the similarity measure proposed in [19], named $XVBr\text{-}\alpha$, which uses a new method to quantify what is understood as a qualitative difference between two IFS elements. A configuration of this measure seem to reflect well the perceived similarity among the simulated XBEs and, moreover, it outperforms the other tested similarity measures.

The second category is constituted by the similarity measures including the "notion of complement" in their definitions [4, 6], as well as by some configurations of the original version of the similarity measure proposed in [19], labeled $XVB\text{-}\alpha\text{-}\omega$. These measures seem to be (partially) effective in comparisons between XBEs resulting from particular scenarios.

The last category consists of similarity measures such as the ones proposed in [20] that could not reflect the perceived similarity between XBEs resulting from opposite scenarios.

Despite the results seem to be significant for choosing a proper similarity measure to compare human XBEs represented as IFSs, they need to be interpreted with caution because of the dependency between the simulated XBEs with the learning algorithm and the contexts that were chosen for the simulations. Hence, in the next chapter an open-source software package whereby a researcher can empirically test similarity measures in the IFS framework will be presented.

References

[1] Krassimir T. Atanassov. *Intuitionistic fuzzy sets.* Fuzzy sets and Systems, 20(1):87–96, 1986.

[2] Krassimir T. Atanassov. *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[3] Eulalia Szmidt and Janusz Kacprzyk. *Geometric similarity measures for the intuitionistic fuzzy sets.* In 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13), pages 840–847. Atlantis Press, 2013.

[4] Eulalia Szmidt. *Similarity Measures between Intuitionistic Fuzzy Sets.* In Distances and Similarities in Intuitionistic Fuzzy Sets, volume 307 of *Studies in Fuzziness and Soft Computing*, pages 87–129. Springer International Publishing, 2014.

[5] Amos Tversky. *Features of similarity.* Psychological review, 84(4):327, 1977.

[6] Eulalia Szmidt and Janusz Kacprzyk. *A concept of similarity for intuitionistic fuzzy sets and its use in group decision making.* In IEEE International Conference on Fuzzy Systems, pages 1129–1134, 2004.

[7] Tony Rose, Mark Stevenson, and Miles Whitehead. *The Reuters Corpus Volume 1-from Yesterday's News to Tomorrow's Language Resources.* In LREC, volume 2, pages 827–832, 2002.

[8] David D. Lewis, Yiming Yang, Tony G. Rose, and Fan Li. *RCV1: A new benchmark collection for text categorization research.* The Journal of Machine Learning Research, 5:361–397, 2004.

[9] Thorsten Joachims. *Text categorization with support vector machines: Learning with many relevant features.* Springer, 1998.

[10] Gerard Salton. *The SMART Retrieval System — Experiments in Automatic Document Processing.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1971.

[11] Martin F. Porter. *An algorithm for suffix stripping.* Program, 14(3):130–137, 1980.

[12] Chris Buckley, Gerard Salton, and James Allan. *The Effect of Adding Relevance Information in a Relevance Feedback Environment.* In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '94, pages 292–300, New York, NY, USA, 1994. Springer-Verlag New York, Inc.

[13] Thorsten Joachims. *Making large-Scale SVM Learning Practical.* In B. Schölkopf, C. Burges, and A. Smola, editors, Advances in Kernel Methods - Support Vector Learning, chapter 11, pages 169–184. MIT Press, Cambridge, MA, 1999.

[14] Zeshui Xu. *Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making.* Fuzzy Optimization and Decision Making, 6(2):109–121, 2007.

[15] Zeshui Xu and Ronald R. Yager. *Intuitionistic and interval-valued intuitionistic fuzzy preference relations and their measures of similarity for the evaluation of agreement within a group.* Fuzzy Optimization and Decision Making, 8(2):123–139, 2009.

[16] Yanhong Li, David L. Olson, and Zheng Qin. *Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis.* Pattern Recognition Letters, 28(2):278 – 285, 2007.

[17] Shyi-Ming Chen, Shou-Hsiung Cheng, and Tzu-Chun Lan. *A novel similarity measure between intuitionistic fuzzy sets based on the centroid points of transformed fuzzy numbers with applications to pattern recognition.* Information Sciences, 343–344:15 – 40, 2016.

[18] Hoang Nguyen. *A novel similarity/dissimilarity measure for intuitionistic fuzzy sets and its application in pattern recognition.* Expert Systems with Applications, 45:97 – 107, 2016.

[19] Marcelo Loor and Guy De Tré. *Connotation-differential Prints - Comparing What Is Connoted Through (Fuzzy) Evaluations.* In Proceedings of the International Conference on Fuzzy Computation Theory and Applications, pages 127–136, 2014.

[20] Marcelo Loor and Guy De Tré. *Vector Based Similarity Measure for Intuitionistic Fuzzy Sets.* In Krassimir T Atanassov, Michał Baczyński, Józef Drewniak, Janusz Kacprzyk, Maciej Krawczak, Eulalia Szmidt, Maciej Wygralak, and Sławomir Zadrożny, editors, Modern Approaches in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics : Volume I: Foundations, pages 105–127. SRI-PAS, 2014.

Chapter 5

IFSMetrics

A Software Package for Studying Similarity Among Intuitionistic Fuzzy Sets

Abstract

Since intuitionistic fuzzy sets (IFSs) have been applied to solve problems in topics like decision-making or pattern recognition, the study of similarity measures aiming to compare this kind of fuzzy sets has become a challenging research subject. When proposed, a similarity measure is usually tested to demonstrate its properties and advantages over the others. However, those tests are in a lot of cases performed using small data sets that do not allow a researcher or practitioner to detect potential drawbacks. As a sequel to the previous chapter, a novel open-source software package, named *IFSMetrics*, by which a researcher can empirically assess several (configurations of) similarity measures while comparing IFSs that characterize experience-based evaluations (XBEs) is proposed in this chapter. By means of this package, one can (1) build a large number of IFSs according to different learning scenarios, (2) compare those IFSs using existing or novel similarity measures, and (3) generate a comprehensive report about how each similarity measure reflects a perceived similarity. Reports generated by the package show that only a few of the existing similarity measures reflect properly a perceived similarity when IFSs resulting from opposite learning scenarios are compared to each other.

This chapter is an adapted version of the following publication:

- Marcelo Loor and Guy De Tré. *An Open-Source Software Package to Assess Similarity Measures that Compare Intuitionistic Fuzzy Sets*. *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Naples, Italy, 2017. This paper was one of the Best Student Papers selected for admission to the Doctoral Consortium at the Fuzz-IEEE 2017 Conference. http://fuzzieee2017.org/doctoral.html.

5.1 Introduction

A similarity comparison is a process where two objects are evaluated to determine the level to which they look like each other. The result of such an evaluation can be quantified through a measure that indicates the *degree of similarity* between those objects. When the objects to compare are (Atanassov) intuitionistic fuzzy sets, IFSs for short [1, 2], a challenge is to create a (similarity) measure that reflects in a proper way the perceived similarity between any two of them.

Several studies aiming to address that challenge in topics like decision-making, e.g., [3, 4, 5] or pattern recognition, e.g., [6, 7, 8], can be found in the literature. In some of those studies, the advantages and properties like *symmetry* or *transitivity* of the proposed similarity measures (see Section 3.4) are tested through a limited number of theoretical or practical examples. A problem that could occur in such cases is that one might not be aware of situations like the described in Section 3.1 in which a similarity comparison between two IFSs characterizing experience-based evaluations (XBEs) yields the highest similarity level although the providers of these XBEs have different understandings of the evaluated concept.

To deal with that problem, in this chapter we propose an open-source software package by which one or more similarity measures can be tested with a large number of IFSs, each characterizing a collection of XBEs resulting from a particular learning scenario. As depicted in Figure 5.1, the proposed package, named *IFSMetrics*, is constituted by three independent modules: *IFSBuilder*, *IFSComparer* and *IFSSimReporter*. While IFSBuilder makes it easier for a researcher or practitioner to extract and build IFSs from a real-world dataset according to multiple scenarios, IFSComparer is the module for comparing those IFSs (or external ones) to each other using the similarity measures therein implemented. The results obtained after comparing such IFSs can be graphically presented using the IFSSimReporter. Herewith, our aim is to provide an integrated software solution that allows a researcher or practitioner to perform adequately controlled experiments with existing or novel similarity measures.

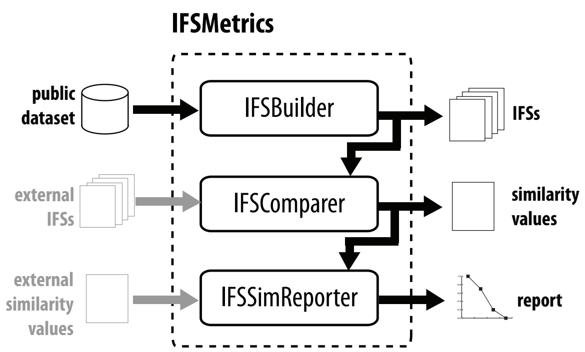


Figure 5.1: A general view of the IFSMetrics package.

A significant aspect of the proposed software package is that, by running controlled experiments with it, one can verify if a particular (configuration of a) similarity measure fits for the purpose of comparing IFSs – e.g., a practitioner can verify if a similarity measure is suitable to compare IFSs representing evaluations with considerable hesitation. Another important feature is that one is free to modify the source code of the package to add new functionalities – e.g., a researcher can include the code of a novel similarity measure to check its usability.

Along with the description of IFSMetrics, which will be presented in Section 5.3, another contribution of this work is the implementation of (almost all of the variants of) the similarity measures proposed in [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]. The source code, which is released under *Apache License 2.0*,[1] can be found in [18]. These contributions are deemed to be significant since they can help a researcher with the assessment of the usability of the similarity measures. Before the description of the package, we present some related work in the next section.

5.2 Related work

In the literature, one can find several efforts that include open-source software tools to support the design and use of systems making use of fuzzy sets which are also called "fuzzy" systems. A comprehensive survey about the current trends on this topic can be found in [19].

Aiming to find any related effort supporting the design and use of similarity measures to compare IFSs, we applied a procedure analogous to the one presented in [19] and introduced the following query into the *Advanced Search* option of the *ISI Web of Knowledge*:

```
TS=( toolbox* OR code* OR software*) AND
TS=(assessment* OR quality OR evaluation OR
validation OR benchmark) AND TS =(similarity*
OR measure*) AND TS = (atanassov* OR fuzzy*
OR intuitionistic*)
```

However, according to the query results (and to the best of our knowledge) such an effort has not been published so far.

Regarding works oriented to determine the suitability of a similarity measure when comparing two IFSs, in [6] several similarity measures were theoretically analyzed according to scenarios that might occur in pattern recognition. In [13, 15, 20] an analogous analysis was performed according to scenarios that might occur in decision-making. In Chapter 4 we designed and applied an empirical procedure to evaluate how suitable a similarity measure is when comparing IFSs that characterize experience-based evaluations. As will be shown in the next section, that procedure has been implemented in the proposed software package to make it easier for a researcher or practitioner to replicate and extend such evaluations.

[1] https://www.apache.org/licenses/LICENSE-2.0

5.3 IFSMetrics

As was mentioned in the introduction to this chapter, *IFSMetrics* is an open-source software package containing modules by which one can build (or derive) IFSs from real-world data, perform similarity comparisons between those IFSs and generate a comprehensive report on the results of such comparisons. As such, *IFSMetrics* is mainly intended for researchers and practitioners, who are deemed to be the *users* of the system. However, it can be used by students, e.g., to observe how a similarity measure reflects what they perceive as similar when two IFSs are compared to each other.

The current version of *IFSMetrics* [18] is constituted by three independent modules: *IFSBuilder*, *IFSComparer* and *IFSSimReporter*. These modules have been written in C#, which is an object-oriented programming language that enables the development of applications in multiple computing platforms. The architecture, main features and functionalities of each module are described next.

5.3.1 IFSBuilder

Through this module a user can build IFSs from real-world data. The current version of *IFSBuilder* is a console application that implements the procedure described in Chapter 4. In that procedure, an IFS represents a collection of XBEs that result from evaluating to which level newswire stories belong to a category. Before conducting such evaluations, a process to learn how to categorize newswire stories according to several learning scenarios is carried out. To perform both the learning and the evaluation processes, the procedure uses the collection of manually categorized newswire stories included in the corrected edition [21] of the dataset provided by Reuter Ltd [22] as ground truth.

To simulate a particular learning scenario, a given percentage of opposite examples in relation to the original data is included in the training collection associated to the scenario. An opposite example consists in a story that will be assigned to a category in the training data when the story has not been assigned to that category in the original data and vice versa. A key aspect on deriving IFSs in this way is that one can observe how a similarity measure reflects a perceived similarity between IFSs resulting from different scenarios taking the ground truth into account. For instance, since one can anticipate that the similarity between IFSs resulting from completely opposite scenarios will be the lowest, one might expect that a similarity measure comparing those IFSs computes the lowest value.

The basic course of actions followed by the building process implemented in *IFSBuilder* is depicted in Figure 5.2. As noticed, an instance of a building process starts when a user configures the scenarios, categories and parameters that will be used during the process. After that, the system, i.e., *IFSBuilder*, verifies the configuration. When the configuration is correct, the system performs both a learning and an evaluation processes for each scenario and for each category.

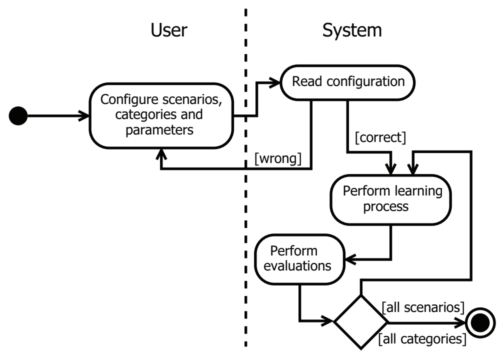


Figure 5.2: Activity diagram of the building process.

The aforementioned course of actions is supported by the general architecture of *IFSBuilder* depicted in Figure 5.3. Notice that to perform a learning process for a given configuration, *IFSBuilder* uses *SVMLight* [23] which implements the *support vector machine learning process* presented in [24, 25]. Using the scenarios and categories specified in the configuration file (see an example of the content of this file in Listing 5.1), the learning process produces a collection of knowledge models. Each model represents the knowledge acquired after learning (from the training collection related to a scenario) what makes newswire stories members of a category. Along with the configuration and the test collection, these models are the inputs of the evaluation process. *IFSBuilder* performs this process to evaluate the stories in the test collection and, also, to characterize the results as IFSs.

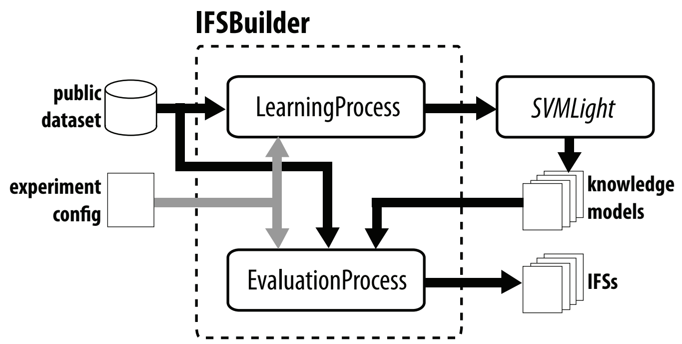


Figure 5.3: A general view of the IFSBuilder module.

Listing 5.1: Experiment configuration.

```
<ExperimentConfig>
  <BuildingProcessParams>
```

```
    <LearningParams/>
    <EvaluationParams>
      <nDocumentsInBatch>50</nDocumentsInBatch>
    </EvaluationParams>
    <Scenarios>
      <Scenario>
        <Code>R20</Code>
        <OppositesPercentage>0.2</OppositesPercentage>
      </Scenario>
    </Scenarios>
    <Categories>
      <Category>
        <Code>ECAT</Code>
      </Category>
    </Categories>
  </BuildingProcessParams>
  <ComparisonProcessParams/>
</ExperimentConfig>
```

It is worth mentioning that each of the returned IFSs results from the evaluations of a fixed number of newswire stories (or documents) specified in an *evaluation batch*. This means that, to obtain an element of an IFS, *IFSBuilder* tests the level to which the proposition p :'x is member of C' is true, where x represents a document in an evaluation batch and C denotes the category under evaluation – recall the definition of IFS presented in Section 3.2. Given that multiple categories and learning scenarios have been considered to build IFSs, the notation $C_{@LS}(X_k)$ was proposed in [12] to denote an IFS resulting after evaluating the proposition p with every document x of X_k according to the learning scenario LS, where X_k represents the collection of documents in the evaluation batch k. Accordingly, *IFSBuilder* uses the notation *'LS-C-SRC-k.ifs'* to name the file that contains the IFS $C_{@LS}(X_k)$ – here *SRC* is the identifier of a test collection. We will explain how to compare those IFSs while describing the *IFSComparer* module in the next subsection.

5.3.2 IFSComparer

By means of this module a user can perform similarity comparisons between IFSs. The current version of *IFSComparer* is a console application in which the similarity measures listed in Table 5.1 have been implemented. A comparison process in this version works as follows:

Consider the collection of newswire stories X. Let X_k be a subset of X. Let also $C_{@LS_i}(X_k)$ and $C_{@LS_j}(X_k)$ be two IFSs resulting from evaluating the proposition p :'x is member of the category C' with each story $x \in X_k$ according to the learning scenarios LS_i and LS_j respectively. Finally, let S be a similarity measure. With these understandings, the comparison process implemented in *IFSComparer* computes the result of $S(C_{@LS_i}(X_k), C_{@LS_j}(X_k))$ for each couple $\langle LS_i, LS_j \rangle$, each category C, each collection of documents X_k and each similarity measure S specified in a given configuration.

Table 5.1: Similarity measures implemented in *IFSComparer*.

Similarity Measure(s)	Proposed in
BA	[17]
CC	[7]
Ch	[9]
Cosine	[15]
Euclidean, Hamming	[14]
GGDM	[4]
HK	[10]
HY15, HY16, HY17	[16]
LOQ	[6]
N26	[8]
SK1, SK2, SK3, SK4	[3]
VB	[11]
Xu17, Xu19, Xu21	[4]
XVB, XVBr	[12]
XY19	[5]

The basic flow of actions followed when a comparison process is performed with *IFSComparer* is shown in Figure 5.4. As noticed, an instance of the comparison process starts after a user configures the scenarios, categories and similarity measures that will be used during the process. After verifying the configuration, the system (i.e., *IFSComparer*) performs a comparison for each of the scenarios, categories and similarity measures established in the configuration file.

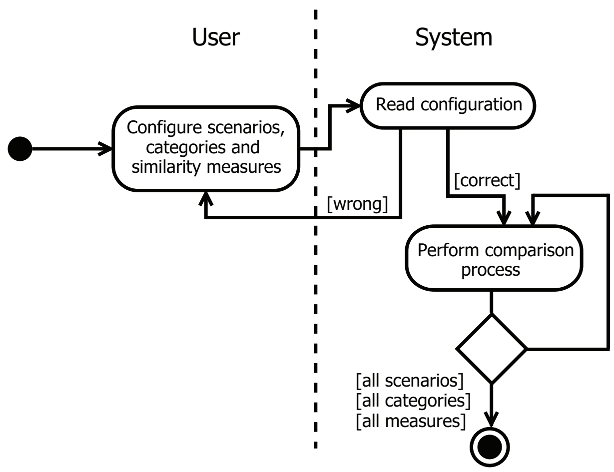


Figure 5.4: Activity diagram of the comparison process.

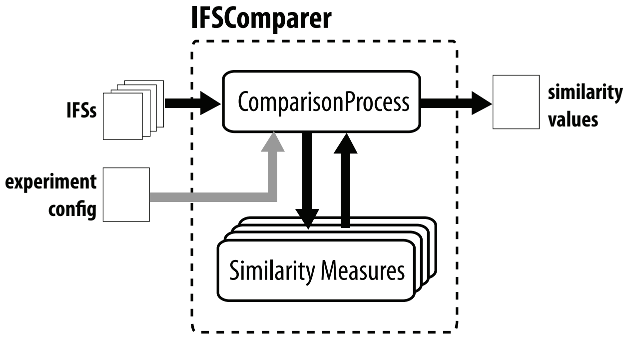


Figure 5.5: A general view of the IFSComparer module.

The general architecture of *IFSComparer* that supports the previous flow of actions is shown in Figure 5.5. Notice that *IFSComparer* can use multiple instances of one or more similarity measures to compute *similarity values* – here, by '*similarity value*' is meant a value in $[0, 1]$ (where 0 denotes not similar at all and 1 means completely similar) that represents the degree of similarity between two IFSs. An important aspect of this design is that a user can test several configurations of a particular similarity measure during the same comparison process. For instance, Listing 5.2 shows the content of an experiment-config file with two configurations of the similarity measure Xu17 [4]: one with $\alpha = 2$ and the other with $\alpha = 0.5$.

Another important feature of the design of *IFSComparer* is that it allows a user to specify a *referent scenario*. By doing so, one can indicate which IFSs will be used as points of reference during a comparison process (which is necessary to test *directional similarity measures*). Thus, one can study how a similarity measure reflects a perceived similarity when IFSs resulting from the referent scenario are compared to IFSs resulting from other scenarios. For example, a user can choose a learning scenario without opposite examples as referent to determine the level to which a similarity measure reflects a perceived similarity between IFSs resulting from this scenario and IFSs resulting from scenarios with opposite examples. In the next section, we will describe how to generate reports that help with this kind of studies.

Listing 5.2: Several configurations of a similarity measure.

```
<ExperimentConfig>
  <BuildingProcessParams/>
  <ComparisonProcessParams>
    <Scenarios/>
    <Categories/>
    <ReferentScenario>
      <Code>R0</Code>
      <OppositesPercentage>0</OppositesPercentage>
    </ReferentScenario>
```

```
    <Measures>
      <Xu17SM>
        <Code>X17-2</Code>
        <Alpha>2</Alpha>
      </Xu17SM>
      <Xu17SM>
        <Code>X17-0.5</Code>
        <Alpha>0.5</Alpha>
      </Xu17SM>
    </Measures>
  </ComparisonProcessParams>
  <ReportingProcessParams/>
</ExperimentConfig>
```

5.3.3 IFSSimReporter

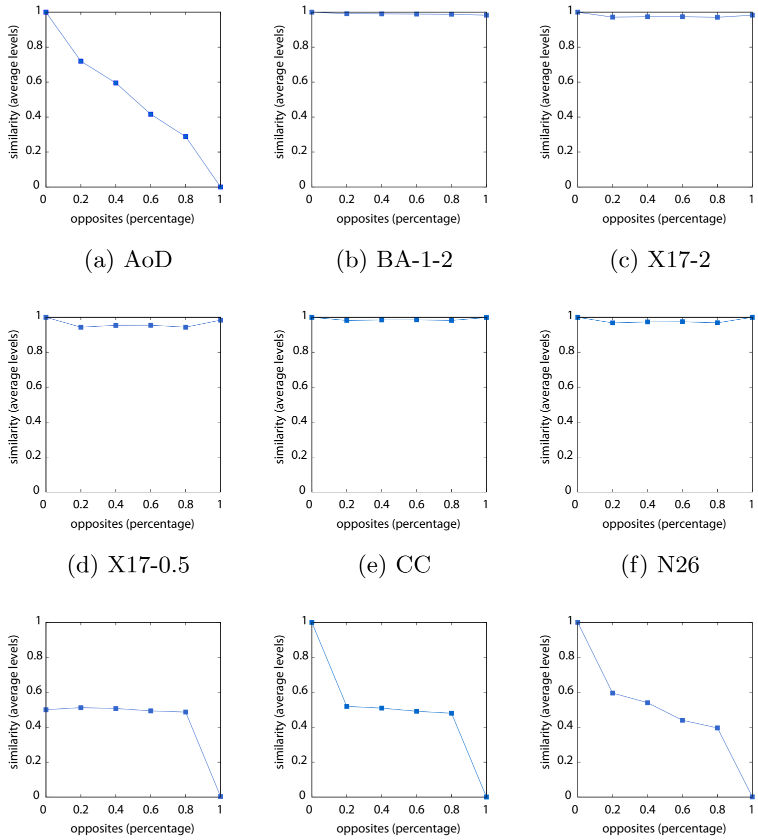
Through the *IFSSimReporter* module a user can build a report that helps with the assessment of the similarity measures used during a comparison process. The current version of *IFSSimReporter* is a console application in which the testing procedure described in [12] has been implemented. In that procedure the following intuition was used to test the similarity measures:

Consider a collection of newswire stories X and a collection $X_k \subseteq X$. Consider also a category C and three learning scenarios, say $R0$, $R20$ and $R100$, established to learn how to categorize newswire stories into C. Consider finally that, while $R0$ is a scenario without opposite examples in the training collection, $R20$ and $R100$ include 20% and 100% of opposite examples respectively. Let $C_{@R0}(X_k)$, $C_{@R20}(X_k)$ and $C_{@R100}(X_k)$ be three IFSs resulting after evaluating the stories in X_k according to knowledge acquired under scenarios $R0$, $R20$ and $R100$ respectively. In this context, one can expect that the level of similarity between $C_{@R0}(X_k)$ and $C_{@R20}(X_k)$ will be greater than or equal to the level of similarity between $C_{@R0}(X_k)$ and $C_{@R100}(X_k)$. One can also expect that the latter level will be the lowest.

By means of the *IFSSimReporter*, one can build a graphical report to illustrate how each of the configured similarity measures reflects the above intuition. For instance, Figure 5.6 was built through *IFSSimReporter*. Notice in this figure that similarity measures like the ones depicted in Figures 5.6b, 5.6c, 5.6d, 5.6e and 5.6f do not reflect the lowest similarity level when IFSs resulting from completely opposite learning scenarios are compared to each other. This means that the similarity levels computed with the aforementioned similarity measures are, in average, in disagreement with the *agreement on decision ratio* (AoD), which is an indicator of the perceived similarity defined in [12].

The basic course of actions followed by *IFSSimReporter* is depicted in Figure 5.7 and the architecture that supports those actions is illustrated in Figure 5.8. Notice that after verifying a proper configuration, the system performs a reporting process for each of the similarity measures included in the similarity-values file. Notice also that *IFSSimReporter* can make use of *Gnuplot*[2] or *pdf-*

[2] http://www.gnuplot.info



(a) AoD (b) BA-1-2 (c) X17-2

(d) X17-0.5 (e) CC (f) N26

(g) XY19 (h) SK1-2D (i) XVBr-0.5-5

Figure 5.6: Averages of the similarity levels per scenario versus the percentage of opposites included in each scenario.

TeX[3] to build a report or parts of it. This is an important characteristic of the design of *IFSSimReporter* since it can provide a researcher with graphics (or tables) that can be easily included in other reports. For instance, a researcher can use those graphics to show how a novel similarity measure outperforms the existing ones when these measures are used to compare IFSs characterizing experience-based evaluations.

It is worth mentioning that, although *IFSSimReporter* and the other modules in *IFSMetrics* have been designed to be independently executed, one can execute them as a whole using the same configuration file. Thus, one can build IFSs, compare them with some similarity measures, and obtain a summarized report of the results all at once. However, since the building process can take a long time to be completed, it is recommended to execute this process only

[3] https://www.tug.org/applications/pdftex/

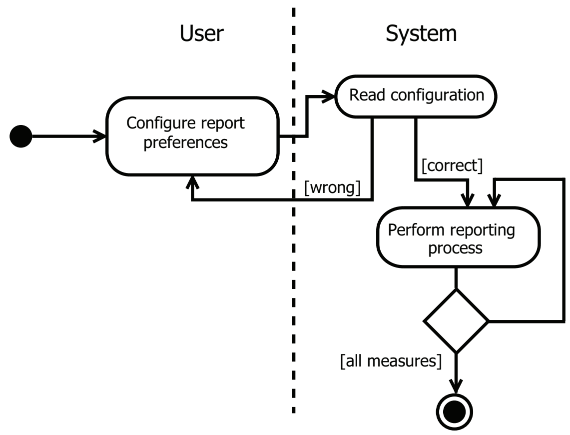


Figure 5.7: Activity diagram of the reporting process.

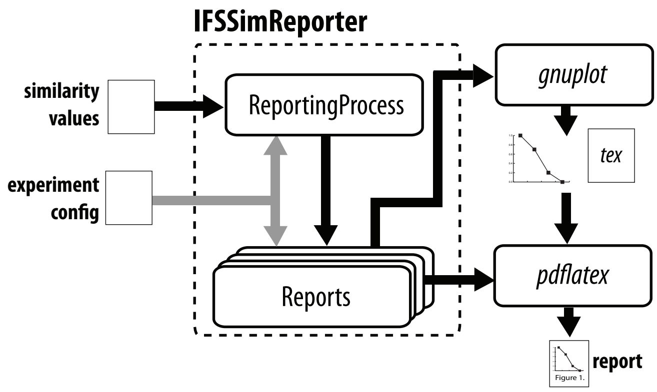


Figure 5.8: A general view of the IFSSimReporter module.

when it is needed. For example, when someone is debugging or looking for a proper configuration of a similarity measure, it is recommended to execute both *IFSComparer* and *IFSSimReporter* as many times as needed to fix the error or find the most suitable configuration, but *IFSBuilder* should be executed only once.

5.4 Conclusions

In this chapter, we have proposed and briefly described a novel open-source software package, named *IFSMetrics*, by which a researcher or practitioner can empirically assess one or more (configurations of) similarity measures that compare intuitionistic fuzzy sets (IFSs). This software package has been developed to support the empirical study presented in Chapter 4. Hence it helps to answer Research Questions $Q2$, $Q3$, $Q4$ and $Q5$.

The current version of the proposed package is composed of three independent modules: *IFSBuilder*, *IFSComparer* and *IFSSimReporter*. By means of *IFSBuilder* one can build a big number of IFSs characterizing experience-based evaluations according to multiple scenarios. Those IFSs can then be compared to each other through the *IFSComparer* module using the similarity measures therein implemented. After that, the results of those comparisons can be processed by *IFSSimReporter* to obtain a comprehensive graphical report about the properties and capabilities of each similarity measure.

By studying a report generated through the package, one can determine the level to which one or more similarity measures reflect what is perceived as similar when IFSs resulting from a referent scenario are compared to IFSs from other scenarios. Thus, a researcher can verify if a particular similarity measure is suited for comparing IFSs in the study under consideration. For instance, one can observe in reports generated by the package that some of the existing similarity measures do not reflect properly a perceive similarity when IFSs resulting from opposite scenarios are compared to each other.

An important feature of the proposed package is that one can modify its source code to add a new similarity measure or fix an existing one. Hence, *IFSMetrics* can be used by a researcher to detect hidden issues that the design of a novel similarity measure could have. In this regard, *IFSMetrics* aims for the design of more reliable similarity measures, which is recommended and subject to further study. This is why the description of this package and the implementation of several similarity measures proposed in the literature are deemed to be significant contributions of this work.

References

[1] Krassimir T. Atanassov. *Intuitionistic fuzzy sets.* Fuzzy sets and Systems, 20(1):87–96, 1986.

[2] Krassimir T. Atanassov. *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[3] Eulalia Szmidt and Janusz Kacprzyk. *A concept of similarity for intuitionistic fuzzy sets and its use in group decision making.* In IEEE International Conference on Fuzzy Systems, pages 1129–1134, 2004.

[4] Zeshui Xu. *Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making.* Fuzzy Optimization and Decision Making, 6(2):109–121, 2007.

[5] Zeshui Xu and Ronald R. Yager. *Intuitionistic and interval-valued intutionistic fuzzy preference relations and their measures of similarity for the evaluation of agreement within a group.* Fuzzy Optimization and Decision Making, 8(2):123–139, 2009.

[6] Yanhong Li, David L. Olson, and Zheng Qin. *Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis.* Pattern Recognition Letters, 28(2):278 – 285, 2007.

[7] Shyi-Ming Chen, Shou-Hsiung Cheng, and Tzu-Chun Lan. *A novel similarity measure between intuitionistic fuzzy sets based on the centroid points of transformed fuzzy numbers with applications to pattern recognition.* Information Sciences, 343–344:15 – 40, 2016.

[8] Hoang Nguyen. *A novel similarity/dissimilarity measure for intuitionistic fuzzy sets and its application in pattern recognition.* Expert Systems with Applications, 45:97 – 107, 2016.

[9] Shyi-Ming Chen. *Similarity measures between vague sets and between elements.* IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 27(1):153–158, 1997.

[10] Dug Hun Hong and Chul Kim. *A note on similarity measures between vague sets and between elements.* Information Sciences, 115(1):83 – 96, 1999.

[11] Marcelo Loor and Guy De Tré. *Vector Based Similarity Measure for Intuitionistic Fuzzy Sets.* In Krassimir T Atanassov, Michał Baczyński, Józef Drewniak, Janusz Kacprzyk, Maciej Krawczak, Eulalia Szmidt, Maciej Wygralak, and Sławomir Zadrożny, editors, Modern Approaches in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics : Volume I: Foundations, pages 105–127. SRI-PAS, 2014.

[12] Marcelo Loor and Guy De Tré. *In a Quest for Suitable Similarity Measures to Compare Experience-Based Evaluations.* In Juan Julián Merelo, Agostinho Rosa, José M. Cadenas, António Dourado Correia, Kurosh Madani, António Ruano, and Joaquim Filipe, editors, Computational Intelligence: International Joint Conference, IJCCI 2015 Lisbon, Portugal, November 12-14, 2015, Revised Selected Papers, volume 669 of *Studies in Computational Intelligence*, pages 291–314. Springer International Publishing, 2017.

[13] Eulalia Szmidt. *Similarity Measures between Intuitionistic Fuzzy Sets.* In Distances and Similarities in Intuitionistic Fuzzy Sets, volume 307 of *Studies in Fuzziness and Soft Computing*, pages 87–129. Springer International Publishing, 2014.

[14] Eulalia Szmidt and Janusz Kacprzyk. *Distances between intuitionistic fuzzy sets.* Fuzzy Sets and Systems, 114(3):505–518, September 2000.

[15] Eulalia Szmidt and Janusz Kacprzyk. *Geometric similarity measures for the intuitionistic fuzzy sets.* In 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13), pages 840–847. Atlantis Press, 2013.

[16] Wen-Liang Hung and Miin-Shen Yang. *Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance.* Pattern Recognition Letters, 25(14):1603 – 1611, 2004.

[17] Fatih Emre Boran and Diyar Akay. *A biparametric similarity measure on intuitionistic fuzzy sets with applications to pattern recognition.* Information Sciences, 255:45 – 57, 2014.

[18] Marcelo Loor and Guy De Tré. *IFSMetrics source code.* https://www.github.com/ifsmetrics-lab/ifsmetrics.

[19] Jesús Alcalá-Fdez and Jose M. Alonso. *A Survey of Fuzzy Systems Software: Taxonomy, Current Research Trends, and Prospects.* IEEE Transactions on Fuzzy Systems, 24(1):40–56, Feb 2016.

[20] Eulalia Szmidt and Janusz Kacprzyk. *A New Similarity Measure for Intuitionistic Fuzzy Sets: Straightforward Approaches may not work.* In 2007 IEEE International Fuzzy Systems Conference, pages 1–6, July 2007.

[21] David D. Lewis, Yiming Yang, Tony G. Rose, and Fan Li. *RCV1: A new benchmark collection for text categorization research.* The Journal of Machine Learning Research, 5:361–397, 2004.

[22] Tony Rose, Mark Stevenson, and Miles Whitehead. *The Reuters Corpus Volume 1-from Yesterday's News to Tomorrow's Language Resources.* In LREC, volume 2, pages 827–832, 2002.

[23] Thorsten Joachims. *Making large-Scale SVM Learning Practical.* In B. Schölkopf, C. Burges, and A. Smola, editors, Advances in Kernel Methods - Support Vector Learning, chapter 11, pages 169–184. MIT Press, Cambridge, MA, 1999.

[24] Vladimir N. Vapnik. *The nature of statistical learning theory.* Springer-Verlag New York, Inc., 1995.

[25] Vladimir N. Vapnik and Vlamimir Vapnik. *Statistical learning theory*, volume 1. Wiley New York, 1998.

Chapter 6

Handling Experience-Based Evaluations with Augmented Appraisal Degrees

Abstract

In Chapters 3, 4 and 5, techniques for handling and comparing experience-based evaluations (XBEs) characterized as elements of intuitionistic fuzzy sets (IFSs) have been studied. Due to the way in which an IFS is defined, such techniques only take into account the magnitude of the appraisal in an XBE. However, as was shown in Chapter 2, such XBEs can be fairly subjective and, thus, their comparison could be affected not only by the magnitude of each appraisal, but also by its context – i.e., a comparison between two XBEs could also be affected by conditions that arise when each evaluation is carried out, which mainly depend on the experience of each evaluator about the topic under consideration. To characterize in a better way the connotative meaning in each XBE, in this chapter an *augmented appraisal degree*, AAD for short, is proposed as a novel generalization of a membership (or non-membership) degree. Along with the definition of an AAD, an augmented framework for handling and comparing XBEs is described. The augmented framework includes several concepts, operators and functions for handling (collections of) AADs. We pay special attention to the description, use, potential benefits and applications of this augmented framework.

This chapter is an adapted version of the following publication:

- Marcelo Loor and Guy De Tré. *On the Need for Augmented Appraisal Degrees to Handle Experience-Based Evaluations. Applied Soft Computing*, 54C (2017): 284-295.

6.1 Introduction

A comparison is a process in which two (or more) objects are examined with the purpose of discovering resemblances or differences. Among others, a comparison is necessary to classify, filter or even arrange objects. Commonly, prior to perform a comparison, an evaluation request is made in order to explain properly what should be taken into account during the process. However, subjective understandings, i.e., understandings affected by a particular experience or knowledge, could lead a comparison to something that differs from the intended request's purpose.

Occasionally, such subjective influence could not be perceived by the requester and, thus, the reliability (and quality) of the results might be affected. If perceived, someone may propose as a solution to formulate a better request or to give a clarification and redo the comparison in order to obtain better results. However, these options may not be practical or feasible in situations such as *citizen science* projects [1], in which a large number of volunteer citizens (with different background) could be asked to evaluate something that cannot be carried out only by a scientific team[1]. Imagine that, in such a situation, the request was already submitted and received a lot of answers, should these all be discarded after reformulating or clarifying it? This problem can be stated as follows: *how to automatically detect and manage any difference in understanding of the purpose behind an evaluation request, in which the answers are given by evaluators with different background.*

In Chapter 3, we studied such a problem considering that each evaluation (or answer) could be imprecise and, moreover, marked by hesitation. First, we modeled an evaluation result as an element of an intuitionistic fuzzy set, IFS for short [6, 7]. Second, we showed that, although comparison strategies for IFSs assuming a metric distance approach work well in many situations (e.g., [8, 9, 10, 11, 12]), they could not handle real-world situations where an evaluation is affected by its context. Third, assuming that a difference in understanding of a concept could be marked by a difference in one or more evaluations of relevant objects, we proposed a *connotation-differential print* (CDP) as a representation of any difference in understandings of a concept in a form that supports further computation. Finally, we presented a straightforward method to build a CDP and used it to augment the results of the similarity measure presented in [13] with the aim of reaching a meaningful comparison between two IFSs. Such an augmented similarity measure helps to overcome an anomaly detected in [14] that happens in some similarity measures in the IFS framework.

In this chapter we propose a novel generalization of an appraisal degree, called *augmented appraisal degree*, within the IFS framework with the main purpose of denoting in a better way the connotative meaning behind each evaluation. Herein, we use the term '*appraisal degree*' to denote the appraisal level of either the *membership* or the *non-membership* of an element in an IFS[2].

[1]Some examples of *citizen science* projects can be found in [2, 3, 4, 5].

[2]Likewise, we shall use the term '*appraisal degree*' to denote: the appraisal level of the *membership* of an element in a *fuzzy set* [15], the appraisal level of either the *membership* or *non-membership* of an element in a *Pythagorean fuzzy set* [16, 17], or even the appraisal

In contrast to what is implicit in the degree-of-similarity semantic interpretation of a membership grade presented in [20], our generalization explicitly takes into account a human behavior in which each person could focus on one or more particular features of an object in order to provide an answer to an evaluation request – this human behavior was studied by Tversky in [21] while presenting his psychological perspective of similarity.

To illustrate the aforementioned human behavior, let us consider the example given in [20], in which one is interested in "*classifying cars*" as 'big cars', 'regular cars' or 'small cars' according to their *known* dimensions or features. When the degree-of-similarity semantic interpretation is used, it is assumed that, to compute the degree of membership of each (object) car to the category 'big', one focuses just on its (feature) size-of-car, which is compared with the (feature) size-of-car from the prototype of a 'big car'. In contrast, our generalization considers that, to compute the degree of membership, additionally to the (feature) size-of-car, some persons might focus on the (feature) size-of-wheels or the (feature) size-of-engine, which are *known* dimensions too. As will be shown in Section 6.2.1, the focus will depend on what features in each car capture the attention of an evaluator according to his/her subjective understanding of 'big cars'. Our motivation here is to find a (mathematical) representation of a membership (or non-membership) grade that reflects in a better way the subjective human behavior described above.

An important and interesting aspect in our approach is that, when the connotative meaning in an XBE is explicitly denoted, dealing with a comparison between two of them becomes a more reliable act. For example, consider the following request: using a unit interval scale where 1 represents the highest level and 0 the lowest, evaluate to which degree a Chihuahua breed is safe for children between 5 and 7. Two evaluators, say A and B, assign 0.6 and 0.7 respectively due to the small size of Chihuahuas; another, say C, assigns 0.6 because their temperamental behavior. If it is assumed that all the evaluators focus on the same features, then the 0.6 from A is equal to the 0.6 from C. If not so, the 0.6@{small-size} from A becomes more similar to the 0.7@{small-size} from B than to the 0.6@{temperamental-behavior} from C. In other words, *the better the connotative meaning in each evaluation is grasped, the more reliable the comparison results are.*

Along with the definition an *augmented appraisal degree* (AAD), which will be presented in Section 6.4, other contributions within this chapter are the following: (a) while answering the question of *how to compare two AADs*, the definition of a *connotation alikeness factor* together with the introduction of the *'as seen from'* and *ℓ-comparison* operators, as well as an *ℓ-difference* function are presented in Section 6.4.1; (b) the definition of an *augmented appraisal function* is given in Section 6.4.2 while answering the question of *how to represent a collection of AADs from a particular point of view*; and (c) an *augmented (Atanassov) intuitionistic fuzzy set* is proposed in Section 6.4.3 while answering the question of *how to handle different points of view within an IFS*. Before the description of these contributions, the psychological perspective

level of either the *satisfaction* or *dissatisfaction* in a *bipolar satisfaction degree* [18, 19].

of similarity under consideration and the formalisms for the representation of appraisal degrees within the framework of IFSs are presented in the next section. Then, some related work and a discussion about their underlying assumptions are presented in Section 6.3.

6.2 Preliminaries

As was mentioned in the introduction to this chapter, a person could focus on one or more features of an object according to his/her individual understandings of the concept behind an evaluation request. Our generalization of a degree-of-similarity semantic of an appraisal degree takes account of such a human behavior. To clarify our motivation for this, a supporting psychological perspective about similarity is given. Hence, the first part of this section presents the psychological perspective proposed by Tversky, in which human behavior in a "feature-matching" based similarity assessment process has been considered [21]. Then, in the second part, the formalisms for representing appraisal degrees within the framework of IFSs are presented.

6.2.1 Similarity according to Tversky

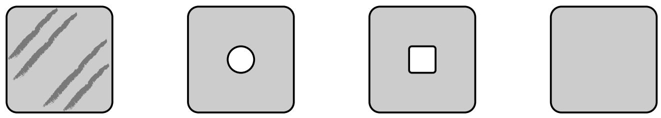
In [21], Tversky proposed a *feature-matching* process to describe the similarity between two objects based on their common and distinctive features. In his approach, Tversky deemed $s(o_1, o_2)$ to be a measure of the *observed (or perceived)* similarity between o_1 and o_2, where $s(o_1, o_2) > s(o_3, o_4)$ indicates that o_1 and o_2 are more similar to each other than o_3 and o_4 are – therein, each object o_i in a referential set U has a collection of features O_i. After that, he proposed a *matching function* $F : U^2 \to \mathbb{R}$ as an approximation of this measure in such a way that $F(o_1, o_2) \geq F(o_3, o_4)$ if and only if $s(o_1, o_2) \geq s(o_3, o_4)$. That approach is based on the following assumptions:

Matching: F depends on the common features (i.e., $O_1 \cap O_2$), the features that belong exclusively to o_1 (i.e., $O_1 - O_2$), and the features that belong exclusively to o_2 (i.e., $O_2 - O_1$). Hence, a measure of the observed similarity between o_1 and o_2 could be denoted as

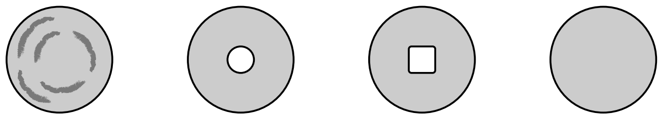
$$s(o_1, o_2) = F(O_1 \cap O_2, O_1 - O_2, O_2 - O_1). \tag{6.1}$$

For example, a common feature between *cookies* c and c' (see Figure 6.1) is the *'square hole'*, a feature that belongs to *cookie* c and not to *cookie* c' is the *'square shape'*, and a feature that belongs to *cookie* c' and not to *cookie* c is the *'round shape'*.

Monotonicity: F increases by adding common features or by decreasing distinctive features or by doing both in conjunction. For instance, putting *'linear icing'* onto the *cookie* d (Figure 6.1d) removes a distinctive feature between it and the *cookie* a (Figure 6.1a), which increases the observed similarity between them.



(a) *cookie a.* (b) *cookie b.* (c) *cookie c.* (d) *cookie d.*



(e) *cookie a'.* (f) *cookie b'.* (g) *cookie c'.* (h) *cookie d'.*

Figure 6.1: Cookies matching (assumptions in a feature-matching process).

Independence: Let A, B, C, D, A', B', C' and D' be collections containing the features present in *cookies* a, b, c, d, a', b', c' and d' (Figure 6.1) respectively[3]. Thus, the common features between *cookie* a and *cookie* b could be denoted by a component $X = A \cap B = \{$*'square shape'*$\}$; the features present in *cookie* a but not in *cookie* b could be denoted by a component $Y = A - B = \{$*'linear icing'*$\}$; and the features present in *cookie* b but not in *cookie* a could be denoted by a component $Z = B - A = \{$*'round hole'*$\}$. Likewise, it could be stated that

$$\begin{aligned}
&A \cap B = A \cap C = \{\text{'square shape'}\} = X,\\
&A' \cap B' = A' \cap C' = \{\text{'round shape'}\} = X',\\
&A - B = A - C = \{\text{'linear icing'}\} = Y,\\
&A' - B' = A' - C' = \{\text{'curved icing'}\} = Y',\\
&B - A = B' - A' = \{\text{'round hole'}\} = Z \text{ and}\\
&C - A = C' - A' = \{\text{'square hole'}\} = Z'.
\end{aligned}$$

Let us imagine that, initially, *cookies* a, b and c, as well as *cookies* a', b' and c', do not have distinctive features, i.e., *cookies* a, b and c all look like *cookie* d, and *cookies* a', b' and c' all look like *cookie* d'. Hence, we might say that the similarity between *cookies* a and b is equal to the similarity between *cookies* a' and b', i.e., $s(a, b) = s(a', b')$. Putting *'linear icing'* onto *cookie* a, as well as a *'curved icing'* onto *cookie* a', may or may not change the order of similarities $s(a, b)$ and $s(a', b')$ regardless of making or not a *'round hole'* in the *cookies* b and b', or making or not a *'square hole'* in the *cookies* c and c'. This means that, when (the components) X and Y, as well as X' and Y' have been taken into account for assessing the similarity, the order of $s(a, b)$ and $s(a', b')$ may change *independently* of Z (or Z').

Solvability: This assumption does not impose constraints on an observed similarity, but just asserts that the corresponding matching function F can be solved.

[3]This example is an adaptation of the example given by Tversky in [21].

Invariance: This assumption neither imposes constraints on an observed similarity, but just states that it is possible to represent similarity by using any interval scale that measures the contribution of $O_1 \cap O_2$, $O_1 - O_2$ and $O_2 - O_1$ in the corresponding matching function F.

Under these assumptions and considering S as an interval similarity scale (or measure) such that $S(o_1, o_2) \geq S(o_3, o_4)$ if and only if $s(o_1, o_2) \geq s(o_3, o_4)$, Tversky then proposed both a *contrast model* and a *ratio model* to represent a matching function F. While in the *contrast model* F has the form

$$S(o_1, o_2) = \lambda_1 \cdot f(O_1 \cap O_2) - \lambda_2 \cdot f(O_1 - O_2) - \lambda_3 \cdot f(O_2 - O_1), \tag{6.2}$$

in the *ratio model* F has the form

$$S(o_1, o_2) = \frac{f(O_1 \cap O_2)}{f(O_1 \cap O_2) + \lambda_2 \cdot f(O_1 - O_2) + \lambda_3 \cdot f(O_2 - O_1)}. \tag{6.3}$$

Herein, λ_1, λ_2 and λ_3 are non-negative numbers, and f is an interval non-negative scale (or measure) of the contribution of a set of features.

As could be noticed, if $\lambda_2 \neq \lambda_3$ and $f(O_1 - O_2) \neq f(O_2 - O1)$, then $S(o_1, o_2) \neq S(o_2, o_1)$ holds in both models, i.e., a matching function could be *asymmetric*. According to Tversky, this reflects what he observed in his experiments regarding the *directionality* of similarity statements, i.e., statements having the form "o_1 is like o_2" – the interested reader is referred to [21] for the description of those experiments.

In those experiments, Tversky also observed that, even though the observed similarity and difference can be considered complementary, it might be influenced by the kind of the evaluation request. When the evaluation request is about the similarity between two objects, an evaluator might pay more attention to the common features; and when the request is about the difference, an evaluator might pay more attention to the distinctive features of the objects. For instance, if the request is about the difference between *cookies* a' and b', we might pay more attention to the *'curved icing'*, which is a feature present only in *cookie* a', and the *'round hole'*, which is a feature present only in *cookie* b', instead of paying attention to the *'round shape'*, which is a common feature between them.

6.2.2 A formalism for representation of appraisal degrees

An XBE could be seen as the result of judging a perceived belongingness (or exclusion) of a subject (i.e., an object that is being evaluated) to a given concept (or category). Within the framework of fuzzy set theory, such a result is always a matter of degree. For instance, one could say "a Chihuahua is a *reasonably* safe breed" instead of just saying "a Chihuahua is a safe breed." In the first case, the object "Chihuahua" is judged as a "reasonable" member of the concept "safe breed," while in the latter case the object is judged as a "full" member of the concept. This means that, when an object is evaluated in this framework, there is room not just for evaluations denoting full belongingness or full exclusion but also for evaluations denoting partial belongingness. In this context, this part presents the essential aspects of appraisal degrees, i.e., membership and non-membership degrees, within the (Atanassov) intuitionistic fuzzy set concept.

6.2.2.1 Membership and non-membership grades

Consider an object x in the universe of discourse X. Consider also a request for the evaluation of the proposition "x is an instance of concept A." In the context of fuzzy set theory [15], an answer to such an evaluation request is not limited to both a *full disagreement* and a *full agreement*, but all the values in between. This means that, e.g., if 0 corresponds to *full disagreement* and 1 corresponds to *full agreement*, an answer could be given by any *intermediate value* between 0 and 1. Here, such a value denotes to which degree x is appraised as an instance (or member) of concept A and, due to this, it is known as a *membership grade* of x in concept A. Moreover, if the membership grade of each $x \in X$ is judged through a function $\mu_A : X \mapsto [0, 1]$, such a function is called a *membership function* of concept A.

As seen above, a membership function could be used to appraise to which degree each object within a group is a member of an individual conception of something known or experienced. Therefore, a membership function could be used to represent the experience-based evaluations given by a particular person about a group of objects. Additionally, considering that in [15] a *fuzzy set* is described by a membership function, those evaluations may be treated as such. In the sequel, as was done in Section 3.2, a fuzzy set A described by a membership function $\mu_A : X \mapsto [0, 1]$ will be denoted by

$$A = \{\langle x, \mu_A(x)\rangle |\, (x \in X) \wedge (0 < \mu_A(x) \leq 1)\}\,.$$

For instance, if $X = \{Chihuahua, Bulldog\}$, A represents a "safe breed", $\mu_A(Chihuahua) = 0.75$ and $\mu_A(Bulldog) = 0.45$, then a fuzzy set that represents an evaluation set for concept A is $\{\langle Chihuahua, 0.75\rangle, \langle Bulldog, 0.45\rangle\}$.

In [6, 7] Atanassov considered that, in addition to a membership grade, a non-membership grade is necessary to appraise whether an object $x \in X$ is an instance or not of concept A. Thus, he proposed an *intuitionistic fuzzy set* (IFS) as an extension of a fuzzy set. An IFS described by a membership function $\mu_A : X \mapsto [0, 1]$ and a non-membership function $\nu_A : X \mapsto [0, 1]$ is denoted by (3.2), i.e.,

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle |\, (x \in X) \wedge (0 \leq \mu_A(x) + \nu_A(x) \leq 1)\}\,.$$

Recall from Section 3.2 that the lack of knowledge about judging an object x as an instance or not of concept A is known as a *hesitation margin* and is expressed by

$$h_A(x_i) = 1 - \mu_A(x_i) - \nu_A(x_i).$$

6.3 Related Work and Discussions

With respect to denoting in a better way the connotative meaning in an evaluation, we found in [22] some ideas of Zadeh about truth and meaning that are related somehow to ours. First, we agree that a person should understand the meaning of a proposition before appraising its truth value. Second, we agree that, although two persons understand the meaning of a proposition, something else is needed to determine the level to which their individual appraisals of the proposition match. Finally, we agree that what is needed is a well-defined (and ready-for-computation) mathematical representation of the meaning of a proposition. However, we consider that a mathematical representation of the connotative meaning in an evaluation of a proposition is needed as well.

Someone may ask: *Is a mathematical representation of the connotative meaning in an evaluation really needed?* To answer this question let us consider the proposition p : '$x = {}^3\!/_{10}$ is similar to ${}^1\!/_3$ according to its value'. This proposition could be mathematically expressed in a canonical form "x IS C" (which means x is an instance of C [23]), where $x = {}^3\!/_{10}$ corresponds to a number in $[0, 1]$ and C corresponds to the (fuzzy) set "numbers in $[0, 1]$ that are similar to ${}^1\!/_3$ according to its value." If two evaluators, say E_1 and E_2, focus on the "metric-distance" and "repeating-decimal" features of (the value of) x respectively, their appraisals of p might be *'more or less true'* and *'quite not true'* in that order. Thus, if the answer from E_2 is expressed without its connotative meaning (i.e., it is just like "p is quite not true"), it could mislead E_1 into thinking that it is wrong. By the contrary, if E_2's answer is expressed like "p is quite not true due to the 'repeating-decimal' feature of x", E_1 can realize that the answer is valid from E_2's perspective. As can be noticed, since the answers (or evaluations) will still depend on the individual understanding of C that each evaluator may have, their comparison will be more reliable with more adequate representations of them. This is the reason why we consider that a mathematical representation of the connotative meaning in an evaluation (or answer) is needed and, therefore, we propose an *augmented appraisal degree* as a generalization of a membership (or non-membership) grade.

6.4 Augmented Appraisal Degrees

Let us begin with an in-a-nutshell description. If a membership (or non-membership) grade denotes to which degree a membership (or non-membership) criterion is fulfilled by an object x, an *augmented appraisal degree* additionally hints *why* such a criterion is fulfilled by x. Here, x represents a person, a notion, or something that exists by itself; and a membership (or non-membership) criterion represents a reason for making a judgment about the membership (or non-membership) of x.

As was pointed out in Section 6.2.1, a person could focus on particular inherent characteristics, i.e., *features*, of an object x in order to make a judgment about it. Hence, we propose to use such features as *hints of a judgment.* This is the basic idea behind the following formal definition:

Definition 6.1 (Augmented appraisal degree (AAD))
Consider a criterion $\mathcal{C}$, an object x and a person P. Let $\mathcal{F}$ be a collection of the features of x. An augmented appraisal degree *of x, say $\hat{\ell}_{\mathcal{C}@P}(x)$, is a pair $\langle \ell_{\mathcal{C}@P}(x), F_{\mathcal{C}@P}(x)\rangle$ that denotes the level $\ell_{\mathcal{C}@P}(x)$ to which x satisfies $\mathcal{C}$, as well as the particular collection of features $F_{\mathcal{C}@P}(x) \subseteq \mathcal{F}$ considered to appraise x from the perspective of P.*

For a better understanding, let us study the following examples of AADs:

1. Consider a number x in $[0, 1]$ and let $\mathcal{C}$ denote the criterion "be a member of A," where A is "the collection of numbers in $[0, 1]$ that are similar to $1/3$ according to their values." Consider also a unit interval scale where 1 and 0 denote, in that order, the highest and the lowest levels of the fulfillment of the criterion. For a particular person, say P, an AAD of $x = 3/10$ satisfying $\mathcal{C}$ can be $\hat{\ell}_{\mathcal{C}@P}(x) = \langle 0.85, \{\textit{metric-distance}\}\rangle$, which indicates that x fulfills $\mathcal{C}$ at 0.85 because of its *metric distance* to $1/3$.
2. Now consider the criterion "do not be a member of A" denoted by $\mathcal{C}'$ and keep A and x the same as above. In this case, an AAD of x satisfying $\mathcal{C}'$ given by another person, say Q, can be $\hat{\ell}_{\mathcal{C}'@Q}(x) = \langle 1\ , \{\textit{repeating-decimal}\}\rangle$, which indicates that x completely fulfills such a *non-membership* criterion due to, in contrast to $1/3$, x is not a *repeating decimal*.

Notice that:

- The criterion in each example is a variant of the form *"membership in A"* or *"non-membership in A"* to highlight how an AAD can be seen as a generalization of a (traditional) membership or non-membership grade in the IFS concept. However, it could also have others forms such as *"compatible with the definition of A," "agreement with A," "disagreement with A," "fittingness in A," "coherence with A,"* etc. Accordingly, we shall use $\hat{\mu}_A$ to denote an AAD of a *membership* criterion, and $\hat{\nu}_A$ to denote a *non-membership* criterion. It is worth mentioning that, instead of using a negation in the definition of a criterion like in the definition of $\mathcal{C}'$, it is preferred to keep the definition like in (the membership criterion) $\mathcal{C}$ and say that $\hat{\nu}_A(x)$ denotes the augmented appraisal degree to which *x dissatisfies $\mathcal{C}$*.
- Even though in the examples A represents a collection, it could also represent a conception of something known, experienced or imagined, i.e., it could represent a *concept*. In this case, a criterion can have a form such as "be compatible with the way in which A is perceived" or "be compatible with the definition of A."
- Notwithstanding in the examples numbers from the unit interval were used for denoting the level to which a criterion is fulfilled, such a the level could also be expressed by a linguistic term like *'largely'* or *'completely'*. In such a case the linguistic term can be modeled by a fuzzy set over the unit interval $[0, 1]$.

- The *metric-distance* and *repeating-decimal* features in the examples show how focusing on different inherent characteristics of the same object x is possible. Even though A should be evaluated on the basis of 'being similar to $1/3$', each evaluator subjectively chooses the final features for the evaluation and these features could be more specific. Surely, the AAD in the second example was given by someone who has experienced at least a bit with repeating decimals. Compare, e.g., with the following descriptions of A: "A is a collection of numbers in $[0, 1]$ that are similar to $1/3$," and "A is a collection of numbers in $[0, 1]$ that are similar to $1/3$ according to the metric distance of their values." In real world user interaction, one can never exclude descriptions that are not specific enough. Hence the motivation for AADs.

6.4.1 Comparing augmented appraisal degrees

As seen in the examples above, two persons, say P and Q, could focus on two different collections of features of an object x, say $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$ respectively, to appraise a criterion $\mathcal{C}$ on x. Moreover, P and Q could choose an appraisal level from different domains, say $\ell\text{-dom}_P$ and $\ell\text{-dom}_Q$ in that order. Thus, if a comparison between two AADs given by P and Q is needed, it is necessary to take both $\ell\text{-dom}_P$-vs.-$\ell\text{-dom}_Q$ and $F_{\mathcal{C}@P}$-vs.-$F_{\mathcal{C}@Q}$ similarities into account. In general, $\ell\text{-dom}_P$ and $\ell\text{-dom}_Q$ are equally fixed for all the evaluators, thus, we can assume hereafter that the appraisal level in any AAD is chosen from the unit interval, i.e., $[0, 1]$. Regarding $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$, as seen before, it is not possible to assume that they are the same. Therefore, to take their similarity into account, we propose a *connotation alikeness factor*, which is defined as follows:

Definition 6.2 (Connotation alikeness factor (CAF))
Consider a criterion $\mathcal{C}$ and two persons, say P and Q. Consider also an object x with a collection of features $\mathcal{F}$. Let $F_{\mathcal{C}@P}, F_{\mathcal{C}@Q} \subseteq \mathcal{F}$ be two collections focused to appraise x from the perspectives of P and Q respectively. A connotation alikeness factor, *CAF for short, is a number $\Delta_{\mathcal{C}:P,Q@R} \in [0, 1]$ that indicates to which level $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$ are similar when $F_{\mathcal{C}@R}$ is taken as a frame of reference. Here, $F_{\mathcal{C}@R}$ could be either $F_{\mathcal{C}@P}$ or $F_{\mathcal{C}@Q}$, and 1 represents the highest level of similarity and 0 the lowest.*

By definition, a CAF will depend on which collection is chosen as a frame of reference, i.e., a CAF is considered to be *directional.* This means that $\Delta_{\mathcal{C}:P,Q@P}$ is not necessarily equal to $\Delta_{\mathcal{C}:P,Q@Q}$ and, thus, a comparison between two AADs will depend on the point of view that is taken as a referent. For example, if $F_{\mathcal{C}@P} = \{a, b, c\}$ and $F_{\mathcal{C}@Q} = \{a, d\}$, the CAF from the perspective of P could be $\Delta_{\mathcal{C}:P,Q@P} = 0.333$; on the contrary, the CAF from the perspective of Q could be $\Delta_{\mathcal{C}:P,Q@Q} = 0.667$.

Since a CAF could be seen as an indicator of the *observed (or perceived)* similarity between $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$, it should reflect the assumptions of the *feature-matching* process presented in Section 6.2.1. Before we illustrate so, it is necessary to know how $\hat{\ell}_{\mathcal{C}@P}$ is seen from the perspective of Q. For that purpose, an '*as seen from*' operator is proposed as follows:

Definition 6.3 ('*as seen from*' operator)
Consider a criterion $\mathcal{C}$, an object x with a collection of features $\mathcal{F}$, and two persons, say P and Q. Assume $I_P = I_Q = [0,1]$. Let $\hat{\ell}_{\mathcal{C}@P}$ be an AAD denoting to which degree x fulfills $\mathcal{C}$ as seen from the perspective of P. An operator 'as seen from' *is a mapping*

$$\begin{aligned} \lfloor \cdot \rfloor_{@Q} : \langle I_P, \mathcal{F} \rangle &\to \langle I_Q, \mathcal{F} \rangle \\ \hat{\ell}_{\mathcal{C}@P} &\mapsto \lfloor \hat{\ell}_{\mathcal{C}@P} \rfloor_{@Q}, \end{aligned} \tag{6.4}$$

such that $\lfloor \hat{\ell}_{\mathcal{C}@P} \rfloor_{@Q}$ is an AAD that corresponds to $\hat{\ell}_{\mathcal{C}@P}$ as seen from the perspective of Q. Here, $\lfloor \hat{\ell}_{\mathcal{C}@P} \rfloor_{@Q}$ has the form $\langle \lfloor \ell_{\mathcal{C}@P} \rfloor_{@Q}, \lfloor F_{\mathcal{C}@P} \rfloor_{@Q} \rangle$, where $\lfloor \ell_{\mathcal{C}@P} \rfloor_{@Q}$ and $\lfloor F_{\mathcal{C}@P} \rfloor_{@Q}$ correspond to $\ell_{\mathcal{C}@P}$ and $F_{\mathcal{C}@P}$ respectively as seen from the perspective of Q.

Now, let us think about the following analogy. Consider a multi-dimensional feature space in which each dimension corresponds to a feature in $\mathcal{F}$. Let $\hat{\mathbf{u}}_{@P}$ be a unit vector that represents an axis related to $F_{\mathcal{C}@P} \subseteq \mathcal{F}$. In this context, an AAD, say $\hat{\ell}_{\mathcal{C}@P} = \langle \ell_{\mathcal{C}@P}, F_{\mathcal{C}@P} \rangle$, could be depicted by a *vector* in which $\ell_{\mathcal{C}@P}$ and $F_{\mathcal{C}@P}$ correspond to its *"magnitude"* and *"direction"* respectively – i.e., $\hat{\ell}_{\mathcal{C}@P}$ corresponds to $\ell_{\mathcal{C}@P} \cdot \hat{\mathbf{u}}_{@P}$. Following the analogy, in Figure 6.2 four AADs representing the evaluations of an object x are depicted: $\hat{\ell}_{\mathcal{C}@P}$, $\hat{\ell}_{\mathcal{C}@Q}$, $\hat{\ell}_{\mathcal{C}@R}$ and $\hat{\ell}_{\mathcal{C}@S}$. All of them have the same appraisal level (or *"magnitude"*), i.e., $\ell_{\mathcal{C}@P} = \ell_{\mathcal{C}@Q} = \ell_{\mathcal{C}@R} = \ell_{\mathcal{C}@S}$, but different focused features (or *"directions"*): $F_{\mathcal{C}@P} = \{a, b\}$, $F_{\mathcal{C}@Q} = \{a, b, c\}$, $F_{\mathcal{C}@R} = \{a, d\}$ and $F_{\mathcal{C}@S} = \{e\}$ respectively. This means that, in this analogy, e.g., the CAF $\Delta_{\mathcal{C}:P,Q@P}$ corresponds to $\cos\theta_{P,Q@P}$, where $\theta_{P,Q@P}$ is the angle between the vectors $\hat{\mathbf{u}}_{@P}$ and $\hat{\mathbf{u}}_{@Q}$ as seen from P. Thereby, the assumptions in the *feature-matching* process are reflected by a CAF as follows:

- *Matching:* From the perspective of who judged $\hat{\ell}_{\mathcal{C}@P}$, i.e., P, the observed similarity between $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@j}$ where $j = Q, R, S$, is denoted by a CAF $\Delta_{\mathcal{C}:P,j@P}$. Here, $\Delta_{\mathcal{C}:P,j@P}$ depends on the common features, i.e., $F_{\mathcal{C}@P} \cap F_{\mathcal{C}@j}$, the features that were focused exclusively by P, i.e., $F_{\mathcal{C}@P} - F_{\mathcal{C}@j}$, and the features that were focused exclusively in $\hat{\ell}_{\mathcal{C}@j}$, i.e., $F_{\mathcal{C}@j} - F_{\mathcal{C}@P}$. For example, $\Delta_{\mathcal{C}:P,Q@P}$ depends on $F_{\mathcal{C}@P} \cap F_{\mathcal{C}@Q} = \{a, b\}$, $F_{\mathcal{C}@Q} - F_{\mathcal{C}@P} = \{c\}$ and $F_{\mathcal{C}@P} - F_{\mathcal{C}@Q} = \{\}$; while $\Delta_{\mathcal{C}:P,S@P}$ depends on $F_{\mathcal{C}@P} - F_{\mathcal{C}@S} = \{a, b\}$, $F_{\mathcal{C}@S} - F_{\mathcal{C}@P} = \{e\}$ and $F_{\mathcal{C}@P} \cap F_{\mathcal{C}@S} = \{\}$.
- *Monotonicity:* A CAF increases by adding common features and/or by decreasing distinctive features. For example, if $c \in F_{\mathcal{C}@Q}$ is not considered in $\hat{\ell}_{\mathcal{C}@Q}$, $\Delta_{\mathcal{C}:P,Q@P}$ will increase. In other words, *the more similar the focused features are, the larger a CAF is.* Notice that, since $\Delta_{\mathcal{C}:P,Q@P} < \Delta_{\mathcal{C}:P,R@P} < \Delta_{\mathcal{C}:P,S@P}$, from the perspective of $\hat{\ell}_{\mathcal{C}@P}$ the comparison $\ell_{\mathcal{C}@P} < \lfloor \ell_{\mathcal{C}@Q} \rfloor_{@P} < \lfloor \ell_{\mathcal{C}@R} \rfloor_{@P} < \lfloor \ell_{\mathcal{C}@S} \rfloor_{@P}$ holds even though $\ell_{\mathcal{C}@P} = \ell_{\mathcal{C}@Q} = \ell_{\mathcal{C}@R} = \ell_{\mathcal{C}@S}$.

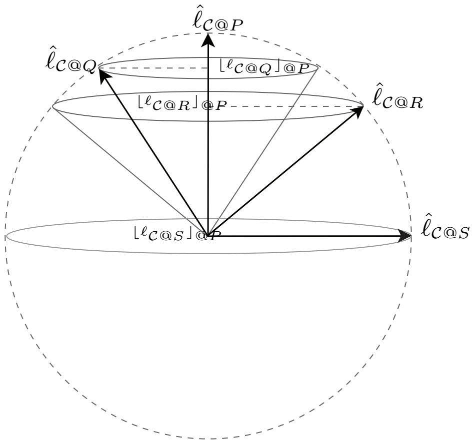


Figure 6.2: AADs depicted as vectors in a multi-dimensional feature space.

- *Independence:* $\Delta_{\mathcal{C}:P,Q@P}$ is independent of $\Delta_{\mathcal{C}:P,j@P}$, where $j = R, S$. That is, a CAF for $\hat{\ell}_{\mathcal{C}@P}$ and $\hat{\ell}_{\mathcal{C}@Q}$ is independent of a CAF for $\hat{\ell}_{\mathcal{C}@P}$ and $\hat{\ell}_{\mathcal{C}@R}$, or a CAF for $\hat{\ell}_{\mathcal{C}@P}$ and $\hat{\ell}_{\mathcal{C}@S}$.
- *Solvability:* The person who judged $\hat{\ell}_{\mathcal{C}@P}$, i.e., P, is able to set $\Delta_{\mathcal{C}:P,Q@P}$ to a particular value because he/she knows the focused features in both $\hat{\ell}_{\mathcal{C}@P}$ and $\hat{\ell}_{\mathcal{C}@Q}$, i.e., $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$. As will be shown in Section 6.4.2.2, if it is not possible for P to know $F_{\mathcal{C}@Q}$, he/she could set $\Delta_{\mathcal{C}:P,Q@P}$ to an approximate value using a *connotation-differential print* [24].
- *Invariance:* $\Delta_{\mathcal{C}:P,Q@P} < \Delta_{\mathcal{C}:P,R@P}$ reflects that the observed similarity between $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$ is less than the observed similarity between $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@R}$.

Returning to the '*as seen from*' operator, from the perspective of Q the features in object x considered for its appraisal are given by $F_{\mathcal{C}@Q}$, thus, it follows that

$$\lfloor F_{\mathcal{C}@P} \rfloor_{@Q} = F_{\mathcal{C}@Q}. \tag{6.5}$$

Moreover, since $I_P = I_Q$, $\lfloor \ell_{\mathcal{C}@P} \rfloor_{@Q}$ will just depend on how similar $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$ are. Therefore, an expression that could be used to obtain its value is

$$\lfloor \ell_{\mathcal{C}@P} \rfloor_{@Q} = \ell_{\mathcal{C}@P} \cdot \Delta_{\mathcal{C}:P,Q@Q}, \tag{6.6}$$

where $\Delta_{\mathcal{C}:P,Q@Q}$ is the CAF for both $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$ that uses $F_{\mathcal{C}@Q}$ as referent. This means that, (6.4) can be denoted by

$$\lfloor \hat{\ell}_{\mathcal{C}@P} \rfloor_{@Q} = \langle \ell_{\mathcal{C}@P} \cdot \Delta_{\mathcal{C}:P,Q@Q}, F_{\mathcal{C}@Q} \rangle. \tag{6.7}$$

For instance, let us rewrite the Tversky's example presented in [21] as follows. Given the criterion "be member of A," where A is a collection of countries that are similar to Russia, two evaluators, say P and Q, state respectively that "Cuba is similar to Russia because of their political affinity" and "Cuba is not similar to Russia because of their geographical proximity." An AAD for P's expression is $\hat{\mu}_{A@P}(Cuba) = \langle 1, \{\textit{political-affinity}\}\rangle$, while another for Q's expression is $\hat{\mu}_{A@Q}(Cuba) = \langle 0, \{\textit{geographical-proximity}\}\rangle$. Using (6.4) from his/her point of view, Q could envisage $\hat{\mu}_{A@P}$ as $\lfloor \hat{\mu}_{A@P} \rfloor_{@Q}$. Thus, if $F_{\mu_A@Q} = \{\textit{geographical-proximity}\}$ and $F_{\mu_A@P} = \{\textit{political-affinity}\}$, then $\lfloor \hat{\mu}_{A@P} \rfloor_{@Q}$ will correspond to $\langle 1 \cdot \Delta_{\mu_A@P,Q@Q}, F_{\mu_A@Q}\rangle$. Moreover, given that $F_{\mu_A@P}$ and $F_{\mu_A@Q}$ have nothing in common, $\Delta_{\mu_A:P,Q@Q}$ could be set to 0 by Q. By doing so, $\lfloor \hat{\mu}_{A@P} \rfloor_{@Q}$ will correspond to $\langle 0, \{\textit{geographical-proximity}\}\rangle$. As might be noticed, if regular membership degrees were used, the evaluations given by P and Q may be fixed to 1 and 0 respectively, which could mislead Q into thinking that P is wrong. By the contrary, using AADs and the '*as seen from*' operator, Q is able to see that P is not wrong, but has a different perspective.

Another benefit of the '*as seen from*' operator is that it can be used to compare two AADs from a particular point of view. For example, a person, say P, could be interested in knowing from his/her perspective if the level in his/her AAD for x and $\mathcal{C}$ is *equal to*, *greater than*, or *less than* the level in an AAD for the same x and the same $\mathcal{C}$ given by someone else. To do so, we define ℓ-comparison operators such as $=_{\ell@P}$, $>_{\ell@P}$, and $<_{\ell@P}$ as follows:

Definition 6.4 (ℓ-comparison operators)
Consider a criterion $\mathcal{C}$ and two persons, say P and Q. Consider also an object x with a collection of features $\mathcal{F}$. Let $\hat{\ell}_{\mathcal{C}@P}$ and $\hat{\ell}_{\mathcal{C}@Q}$ be two AADs denoting to which degree x fulfills $\mathcal{C}$ as seen from the perspectives of P and Q respectively. Assume $I = [0, 1]$. An operator $\mathrm{cmp}_{\ell@P} : \langle I, \mathcal{F}\rangle^2 \to \{0, 1\}$ *is called an* ℓ-comparison operator *if, from the perspective of P, the following holds for $\ell_{\mathcal{C}@P}$ in $\hat{\ell}_{\mathcal{C}@P}$ and $\lfloor \ell_{\mathcal{C}@Q} \rfloor_{@P}$ in $\lfloor \hat{\ell}_{\mathcal{C}@Q} \rfloor_{@P}$:*

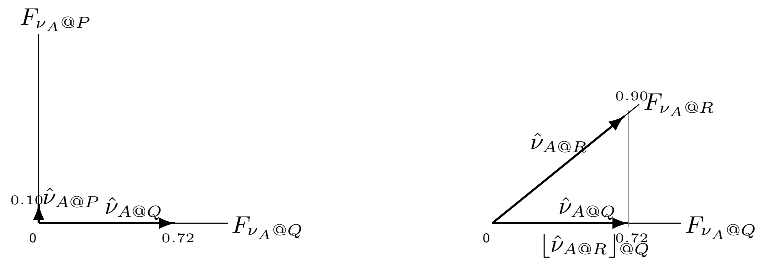
$$\ell_{\mathcal{C}@P} \ \mathrm{cmp} \ \lfloor \ell_{\mathcal{C}@Q} \rfloor_{@P}, \tag{6.8}$$

where cmp *is a* numerical comparison operator *(e.g., $=$, $>$ or $<$).*

By definition, e.g., $=_{\ell@P}$, $>_{\ell@P}$ and $<_{\ell@P}$ are related to the numerical comparison operators $=$, $>$ and $<$ respectively, as follows:

$$\begin{aligned}
&\hat{\ell}_{\mathcal{C}@P} =_{\ell@P} \hat{\ell}_{\mathcal{C}@Q} \text{ if and only if } \ell_{\mathcal{C}@P} = \lfloor \ell_{\mathcal{C}@Q} \rfloor_{@P},\\
&\hat{\ell}_{\mathcal{C}@P} >_{\ell@P} \hat{\ell}_{\mathcal{C}@Q} \text{ if and only if } \ell_{\mathcal{C}@P} > \lfloor \ell_{\mathcal{C}@Q} \rfloor_{@P} \text{ and}\\
&\hat{\ell}_{\mathcal{C}@P} <_{\ell@P} \hat{\ell}_{\mathcal{C}@Q} \text{ if and only if } \ell_{\mathcal{C}@P} < \lfloor \ell_{\mathcal{C}@Q} \rfloor_{@P}.
\end{aligned}$$

For example, for the non-membership-in-A criterion, where A is *"healthy sports"* and the object *"football,"* three evaluators, say P, Q and R, have given their AADs, say $\hat{\nu}_{A@P}$, $\hat{\nu}_{A@Q}$ and $\hat{\nu}_{A@R}$ respectively, as follows: $\langle 0.10, \{\textit{duration-of-the-match}\}\rangle$ from P, $\langle 0.72, \{\textit{physical-contact}, \textit{unsafe-outfit}\}\rangle$ from Q; and $\langle 0.90, \{\textit{physical-contact}\}\rangle$ from R. From his/her perspective, evaluator Q could fix $\Delta_{\nu:P,Q@Q} = 0$ because there is nothing in common between his/her focused features in *"football"* and the focused features from P. If so, using (6.6), the



(a) $\hat{\nu}_{A@Q}$ vs $\hat{\nu}_{A@P}$. (b) $\hat{\nu}_{A@Q}$ vs $\hat{\nu}_{A@R}$

Figure 6.3: An analogy in a comparison between two AADs.

comparison $\hat{\nu}_{A@Q} =_{\ell@Q} \hat{\nu}_{A@P}$ corresponds to $(0.72 = 0.10 \cdot 0)$ and, so, its result is 0 (i.e., *false*). On the other hand, if Q fixes $\Delta_{\nu:Q,R@Q} = 0.80$ because he/she perceives *physical-contact* as an important feature, then the comparison $\hat{\nu}_{A@Q} =_{\ell@Q} \hat{\nu}_{A@R}$ corresponds to the evaluation of $(0.72 = 0.90 \cdot 0.80)$ and, so, its result is 1 (i.e., *true*). Notice that, to fix the values in $\Delta_{\nu:P,Q@Q}$ and $\Delta_{\nu:Q,R@Q}$, Q follows the assumptions of the *feature-matching* process described in Section 6.2.1. This means that, e.g., Q considers that $\Delta_{\nu:P,Q@Q}$ is the result of a function $s(F_{\nu_A@P}, F_{\nu_A@Q})$ such as (6.1), where $F_{\nu_A@P} = \{$*duration-of-the-match*$\}$ and $F_{\nu_A@Q} = \{$*physical-contact*, *unsafe-outfit*$\}$.

To clarify further an ℓ-comparison between two AADs, let us use the analogy written above about depicting an AAD $\hat{\ell}_{\mathcal{C}@P}$ by a vector. Thereby, $\hat{\nu}_{A@P}$, $\hat{\nu}_{A@Q}$ and $\hat{\nu}_{A@R}$ from the previous example could be depicted as shown in Figure 6.3. In Figure 6.3a, since $F_{\nu_A@Q}$ and $F_{\nu_A@P}$ have nothing in common, the *"magnitude"* of $\hat{\nu}_{A@P}$, i.e., $\nu_{A@P} = 0.10$, is seen as $\lfloor \nu_{A@P} \rfloor_{@Q} = 0$ in the *"direction"* of $\hat{\nu}_{A@Q}$ – since in this analogy $\Delta_{\nu:P,Q@Q} = 0$ corresponds to $\cos \theta_{P,Q@Q} = 0$ where $\theta_{P,Q@Q}$ is the angle between the "vectors" $\hat{\nu}_{A@Q}$ and $\hat{\nu}_{A@P}$, they are depicted perpendicularly. On the contrary, in Figure 6.3b, since $F_{\nu_A@Q}$ and $F_{\nu_A@R}$ have an important feature in common, the *"magnitude"* of $\hat{\nu}_{A@R}$, i.e., $\nu_{A@R} = 0.90$ is seen as $\lfloor \nu_{A@R} \rfloor_{@Q} = 0.72$ in the *"direction"* of $\hat{\nu}_{A@Q}$ – this is why the result of $\hat{\nu}_{A@Q} =_{\ell@Q} \hat{\nu}_{A@R}$ holds.

It is worth mentioning that, by using $[0, 1]$ instead $\{0, 1\}$ as a co-domain in Definition 6.4, we can define fuzzy variants such as *'approximately equal'*, *'not much larger than'*, or *'much smaller than'* of the *ℓ-comparison operators*. For example, the result of $\hat{\ell}_{\mathcal{C}@P}$ *'approximately equal'*$_{\ell@P}$ $\hat{\ell}_{\mathcal{C}@Q}$ will depend on the observed difference between $\ell_{\mathcal{C}@P}$ and $\lfloor \ell_{\mathcal{C}@Q} \rfloor_{@P}$ from the perspective of P. To measure such a difference, we propose an *ℓ-difference* mapping, which is defined as follows:

Definition 6.5 (ℓ-difference)
Consider a criterion $\mathcal{C}$ and two persons, say P and Q. Consider also an object x with a collection of features $\mathcal{F}$. Let $\hat{\ell}_{\mathcal{C}@P}$ and $\hat{\ell}_{\mathcal{C}@Q}$ be two AADs denoting to which degree x fulfills $\mathcal{C}$ as seen from the perspectives of P and Q respectively. Assume $I = [0, 1]$. An ℓ-difference *is a mapping* $\mathrm{dif}_{\ell@P} : \langle I, \mathcal{F} \rangle^2 \to [-1, 1]$ *such*

that

$$\mathrm{dif}_{\ell @P}(\hat{\ell}_{\mathcal{C}@P}, \hat{\ell}_{\mathcal{C}@Q}) = \ell_{\mathcal{C}@P} - \lfloor \ell_{\mathcal{C}@Q} \rfloor_{@P} \tag{6.9}$$

is a measure of the difference between $\hat{\ell}_{\mathcal{C}@P}$ *and* $\hat{\ell}_{\mathcal{C}@Q}$ *from the perspective of* P.

A result of an *ℓ-difference* can be interpreted from the perspective of P like this: when $\mathrm{dif}_{\ell @P}(\hat{\ell}_{\mathcal{C}@P}, \hat{\ell}_{\mathcal{C}@Q}) = 0$, there is no difference between $\hat{\ell}_{\mathcal{C}@P}$ and $\hat{\ell}_{\mathcal{C}@Q}$; when $\mathrm{dif}_{\ell @P}(\hat{\ell}_{\mathcal{C}@P}, \hat{\ell}_{\mathcal{C}@Q}) > 0$, the appraisal for x from P is greater than the appraisal given by Q; and when $\mathrm{dif}_{\ell @P}(\hat{\ell}_{\mathcal{C}@P}, \hat{\ell}_{\mathcal{C}@Q}) < 0$, the appraisal for x from P is less than the appraisal given by Q.

Using an *ℓ-difference*, $\hat{\ell}_{\mathcal{C}@P}$ *'approximately equal'*$_{\ell @P}$ $\hat{\ell}_{\mathcal{C}@Q}$ may be interpreted as $F\left(\left|\mathrm{dif}_{\ell @P}(\hat{\ell}_{\mathcal{C}@P}, \hat{\ell}_{\mathcal{C}@Q})\right|\right)$, where $F : [0,1] \mapsto [0,1]$ is the membership function that, according to P, represents the concept *'approximately equal'* [20, 25]. In the same way, $\hat{\ell}_{\mathcal{C}@P}$ *'much smaller than'*$_{\ell @P}$ $\hat{\ell}_{\mathcal{C}@Q}$ may be interpreted as $G\left(\left|\mathrm{dif}_{\ell @P}(\hat{\ell}_{\mathcal{C}@P}, \hat{\ell}_{\mathcal{C}@Q})\right|\right)$, where $G : [0,1] \mapsto [0,1]$ is the membership function that represents the concept *'much smaller than'* from the perspective of P.

6.4.2 Augmented appraisal functions

So far, we explained how an AAD can denote the judgment of an object x fulfilling a criterion $\mathcal{C}$. Moreover, we have seen that a criterion $\mathcal{C}$ has, among others, a form such as "be member of A" or "be compatible with the definition of A," where A can represent a concept or a collection related to a concept. However, there are also cases in which it is needed to judge not one, but many objects within a collection X. In such cases, one can be interested in representing the correspondence between each $x \in X$ and its AAD through a function that denotes his/her point of view about a criterion $\mathcal{C}$. To do so, we propose an *augmented appraisal function*, which is defined as follows:

Definition 6.6 (Augmented appraisal function (AAF))
Consider a criterion $\mathcal{C}$, *a collection of objects* $X = \{x_1, \cdots, x_n\}$, *and a person* P. *Let* $\mathcal{F}_i$ *be the collection of the features of an object* x_i *in* X. *Assume* $I = [0,1]$ *and* $\mathcal{F} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_n$. *An* augmented appraisal function *is a mapping* $\hat{\ell}_{\mathcal{C}@P} : X \to \langle I, \mathcal{F} \rangle$ *that denotes the correspondence between each* $x_i \in X$ *and* $\langle \ell_{\mathcal{C}@P}(x_i), F_{\mathcal{C}@P}(x_i) \rangle$, *where* $\langle \ell_{\mathcal{C}@P}(x_i), F_{\mathcal{C}@P}(x_i) \rangle$ *is an AAD denoting the appraisal of* x_i *fulfilling* $\mathcal{C}$ *as seen from the perspective of* P.

By definition, an AAF reflects *judgments* of each object in X from a particular *point of view.* In a *degree-of-similarity* semantic interpretation of an appraisal degree, one judges an object according to which of its features are similar or not to the features in a *prototype* that represents his/her understanding of a concept, say A. Thereby, given a criterion such as "be compatible with the definition of A" and a collection X, an AAF will reflect *judgments* for each $x_i \in X$ according to how similar its features and the features of a *prototype* for

Table 6.1: Examples of augmented appraisal functions for the criterion *'be compatible with the definition of A,"* where A is defined as a collection of numbers in $[0, 1]$ that are similar to $1/3$.

i	x_i	$\hat{\mu}_{A@P}(x_i)$	$\hat{\mu}_{A@Q}(x_i)$
1	$4/45$	$\langle 0.75, \textit{repeating-decimal}\rangle$	$\langle 0.27, \textit{metric-distance}\rangle$
2	$1/10$	$\langle 0, \textit{repeating-decimal}\rangle$	$\langle 0.30, \textit{metric-distance}\rangle$
3	$1/9$	$\langle 1, \textit{repeating-decimal}\rangle$	$\langle 0.33, \textit{metric-distance}\rangle$
4	$3/10$	$\langle 0, \textit{repeating-decimal}\rangle$	$\langle 0.90, \textit{metric-distance}\rangle$
5	$1/3$	$\langle 1, \textit{repeating-decimal}\rangle$	$\langle 1, \textit{metric-distance}\rangle$
6	$1/2$	$\langle 0, \textit{repeating-decimal}\rangle$	$\langle 0.50, \textit{metric-distance}\rangle$
7	$2/3$	$\langle 1, \textit{repeating-decimal}\rangle$	$\langle 0, \textit{metric-distance}\rangle$

A are. In others words, the *judgments* in an AAF will depend on the *prototype* for a concept given from a particular *point of view.* For instance, consider the criterion *'be compatible with the definition of A,"* where A is defined as a collection of numbers in $[0, 1]$ that are similar to $1/3$. Consider also the collection $X = \{4/45, 1/10, 1/9, 3/10, 1/3, 1/2, 2/3\}$. Table 6.1 shows the AAFs $\hat{\mu}_{A@P}$ and $\hat{\mu}_{A@Q}$ from two persons P and Q respectively, whose prototypes for A are described as follows: for P, the prototype is a number in $[0, 1]$ that consists of a repeating decimal, i.e., that is of the the form '$0.ddd\cdots$' with $d \in \{1, \cdots, 9\}$; and for Q, the prototype is a number $x \in [0, 2/3]$ such that x is as close as possible to $1/3$ according to its metric distance. As could be noticed, $\hat{\mu}_{A@P}(x_i)$ and $\hat{\mu}_{A@Q}(x_i)$ hint for each x_i which of its features are contrasted with the features of their corresponding prototypes.

An alternative form for representing the correspondence between each $x_i \in X$ and its AAD is by means of a set

$$\hat{A}_{@P} = \left\{ \langle x_i, \hat{\ell}_{\mathcal{C}@P}(x_i)\rangle |\, (x_i \in X) \wedge \left(\hat{\ell}_{\mathcal{C}@P}(x_i) \in \langle I, \mathcal{F}\rangle\right)\right\}. \tag{6.10}$$

For example, AAFs $\hat{\mu}_{A@P}$ and $\hat{\mu}_{A@Q}$ from Table 6.1 can be represented as

$$\begin{aligned}\hat{A}_{@P} &= \{\langle x_i, \hat{\mu}_{A@P}(x_i)\rangle |\, (x_i \in X) \wedge (\hat{\mu}_{A@P}(x_i) \in \langle I, \mathcal{F}\rangle)\} \text{ and}\\ \hat{A}_{@Q} &= \{\langle x_i, \hat{\mu}_{A@Q}(x_i)\rangle |\, (x_i \in X) \wedge (\hat{\mu}_{A@Q}(x_i) \in \langle I, \mathcal{F}\rangle)\}\end{aligned}$$

respectively (cf. (3.1)). This form has a main advantage: it makes dealing with collections of XBEs easier – e.g., $\hat{A}_{@P}$ and $\hat{A}_{@Q}$ in Table 6.1 can be treated as two collections consisting of XBEs of "numbers in $[0, 1]$ that are similar to $1/3$" given by P and Q respectively.

6.4.2.1 Comparing augmented appraisal functions

As stated in Definition 6.2, a CAF indicates how similar two collections of features of an object, obtained from two perspectives, are. Considering the *degree-of-similarity* interpretation of an appraisal degree given above, we could also say that *a CAF indicates how similar (the features in) the prototypes from* *two AAFs are.* Therefore, one could use a CAF to figure out how each AAD in an AAF looks like considering another perspective. For example, from P's perspective, the CAF for $\hat{\mu}_{A@P}$ and $\hat{\mu}_{A@Q}$ from Table 6.1 is $\Delta_{\mu:P,Q@P} = 0$ because there is nothing in common between their prototypes; thus, e.g., $\lfloor \hat{\mu}_{A@Q}(1/9) \rfloor_{@P}$ will correspond to $\langle 0, \textit{repeating-decimal} \rangle$. Using this idea, a person could be interested in comparing from his/her perspective two AAFs for a collection X and a criterion $\mathcal{C}$. To do so, we propose the use of an *ℓ-similarity measure*, which is defined as follows:

Definition 6.7 (ℓ-similarity measure)
Consider a criterion $\mathcal{C}$ and two persons, say P and Q. Consider also a collection of objects $X = \{x_1, \cdots, x_n\}$, where each $x_i \in X$ has a collection of features $\mathcal{F}_i$. Assume $I = [0,1]$ and $\mathcal{F} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_n$. Let $\hat{\ell}_{\mathcal{C}@P} : X \to \langle I, \mathcal{F} \rangle$ and $\hat{\ell}_{\mathcal{C}@Q} : X \to \langle I, \mathcal{F} \rangle$ be two AAFs denoting to which degree each $x_i \in X$ fulfills $\mathcal{C}$ as seen from the perspectives of P and Q respectively. Furthermore, assume

$$\hat{A}_{@P} = \left\{ \langle x_i, \hat{\ell}_{\mathcal{C}@P}(x_i) \rangle | \, (x_i \in X) \wedge \left(\hat{\ell}_{\mathcal{C}@P}(x_i) \in \langle I, \mathcal{F} \rangle \right) \right\} \text{ and}$$
$$\hat{A}_{@Q} = \left\{ \langle x_i, \hat{\ell}_{\mathcal{C}@Q}(x_i) \rangle | \, (x_i \in X) \wedge \left(\hat{\ell}_{\mathcal{C}@Q}(x_i) \in \langle I, \mathcal{F} \rangle \right) \right\}.$$

An ℓ-similarity measure, *say* $\mathrm{sim}_{\ell@P}$, *is a measure of the level to which* $\hat{A}_{@P}$ *and* $\hat{A}_{Q}$ *are similar as seen from the perspective of* P.

As might be noticed, the modeling of $\hat{A}_{@P}$ and $\hat{A}_{@Q}$ highlights how a comparison between two AAFs could be seen as a comparison between two fuzzy sets. However, in contrast to what is considered in [26] for similarity relations between two fuzzy sets, an *ℓ-similarity measure* is considered to be *asymmetrical* because it depends on the perspective taken as a point of reference, i.e., $\mathrm{sim}_{\ell@P}$ is not necessarily equal to $\mathrm{sim}_{\ell@Q}$. Therefore, to obtain an expression for $\mathrm{sim}_{\ell@P}$, we should use a model such as the *ratio model* (see Section 6.2.1), which does not assume symmetry in similarity relations. For that purpose, recalling from Section 6.2.1 that the similarity and difference are complementary, we use (6.9) to obtain a difference between $\hat{\ell}_{\mathcal{C}@P}(x_i)$ and $\hat{\ell}_{\mathcal{C}@Q}(x_i)$ for each $x_i \in X$, and then, we aggregate all the differences to obtain

$$\mathrm{sim}_{\ell@P}(\hat{A}_{@P}, \hat{A}_{@Q}) = 1 - \frac{1}{n} \sum_{i=1}^{n} \left| \mathrm{dif}_{\ell@P} \left(\hat{\ell}_{\mathcal{C}@P}(x_i), \hat{\ell}_{\mathcal{C}@Q}(x_i) \right) \right| \tag{6.11}$$

as an *ℓ-similarity measure* of the similarity between $\hat{A}_{@P}$ and $\hat{A}_{@Q}$ as seen from the perspective of P (cf. (3.21)). To show that this expression is conform to the *ratio model* in (6.3), we could rewrite it as

$$\mathrm{sim}_{\ell@P}(\hat{A}_{@P}, \hat{A}_{@Q}) = \frac{g_1(\hat{A}_{@P} \cap \hat{A}_{@Q})}{g_1(\hat{A}_{@P} \cap \hat{A}_{@Q}) + \lambda_2 g_2(\hat{A}_{@P} - \hat{A}_{@Q}) + \lambda_3 g_3(\hat{A}_{@Q} - \hat{A}_{@P})},$$

where

$$
\begin{aligned}
g_2(\hat{A}_{@P} - \hat{A}_{@Q}) &= \begin{cases} \sum_{i=1}^{n} f(x_i) & \forall f(x_i) : f(x_i) > 0 \\ 0 & \text{otherwise,} \end{cases} \\
g_3(\hat{A}_{@Q} - \hat{A}_{@P}) &= \begin{cases} \sum_{i=1}^{n} |f(x_i)| & \forall f(x_i) : f(x_i) < 0 \\ 0 & \text{otherwise,} \end{cases} \\
g_1(\hat{A}_{@P} \cap \hat{A}_{@Q}) &= n - g_2(\hat{A}_{@P} - \hat{A}_{@Q}) - g_3(\hat{A}_{@Q} - \hat{A}_{@P}), \\
f(x_i) &= \mathrm{dif}_{\ell@P}\left(\hat{\ell}_{\mathcal{C}@P}(x_i), \hat{\ell}_{\mathcal{C}@Q}(x_i)\right),
\end{aligned}
$$

and $\lambda_2 = \lambda_3 = 1$.

As an example of using (6.11), let us figure out both $\mathrm{sim}_{\ell@P}(\hat{A}_{@P}, \hat{A}_{@Q})$ and $\mathrm{sim}_{\ell@Q}(\hat{A}_{@P}, \hat{A}_{@Q})$ for the AADs in Table 6.1. As has been shown in the description of the example above, from P's perspective, a CAF for the prototypes from P and Q is $\Delta_{\mu_A:P,Q@P} = 0$. Thus, using (6.9) together with (6.6), we obtain $\mathrm{sim}_{\ell@P}(\hat{A}_{@P}, \hat{A}_{@Q}) = 0.46$. Supposing that Q also sets $\Delta_{\mu_A:P,Q@Q} = 0$, $\mathrm{sim}_{\ell@Q}(\hat{P}, \hat{Q}) = 0.53$ is obtained as well. Notice that $\mathrm{sim}_{\ell@P}(\hat{A}_{@P}, \hat{A}_{@Q})$ is not equal to $\mathrm{sim}_{\ell@Q}(\hat{A}_{@P}, \hat{A}_{@Q})$, which reflects the *asymmetry* in (6.11).

6.4.2.2 Comparing augmented and regular appraisal functions

As was mentioned in the previous section, a comparison between two AAFs depends on a CAF given from a particular perspective. It was also mentioned that such a CAF can be fixed according to the observed similarity between the collections of features focused on in the prototypes for each AAF, which are visible through the constituent AADs. However, there are cases where it is not possible to get two AAFs, but just one – i.e., the constituent AADs are visible only in one AAF. In such cases, we could use a method to approximate the value of a CAF. Next we describe a straightforward method to approximate a CAF based on the idea behind the *connotation-differential print* (CDP) presented in Section 3.3.3.

During the process carried out to evaluate each object $x_i \in X$ fulfilling a criterion $\mathcal{C}$, one could shift his/her focus onto some objects having features that he/she deemed to be relevant because they highly satisfy (or dissatisfy) $\mathcal{C}$. For instance, consider the person whose evaluations are denoted by $\hat{\mu}_{A@P}$ in Table 6.1, i.e., P. P could shift his/her focus onto $1/9$ or $2/3$ because, from his/her perspective, they fully satisfy the criterion "be member of A," where A is a collection of numbers in $[0, 1]$ that are similar to $1/3$. As was mentioned in Section 3.3.3, the idea behind a CDP is that *a difference in connotation of A could be marked by a difference in the appraisal levels given for some relevant objects.*

To denote the relative difference between two appraisal levels given for a relevant object, we could use a representation similar to a *connotation-differential marker* (CDM) (see Definition 3.1). In this case, considering appraisal levels

$\ell_{\mathcal{C}@P}(x_i)$ and $\ell_{\mathcal{C}@Q}(x_i)$ given by P and Q respectively for object x_i, a CDM will be a symbol $s \in \{\diamond\!\!\!\mid, \overset{\diamond}{\uparrow}, \underset{\diamond}{\downarrow}\}$ following the conditions:

$$\begin{aligned}
&\text{if } |\ell_{\mathcal{C}@P}(x_i) - \ell_{\mathcal{C}@Q}(x_i)| \leq \delta \text{ then } s = \diamond\!\!\!\mid,\\
&\text{if } \ell_{\mathcal{C}@P}(x_i) - \ell_{\mathcal{C}@Q}(x_i) > \delta \text{ then } s = \overset{\diamond}{\uparrow},\\
&\text{if } \ell_{\mathcal{C}@P}(x_i) - \ell_{\mathcal{C}@Q}(x_i) < -\delta \text{ then } s = \underset{\diamond}{\downarrow},
\end{aligned}$$

where $\delta \in [0, 1]$.

Accordingly, a CDP will correspond to a sequence of CDMs obtained for the objects considered to be relevant from a particular point of view. For instance, in the example above, P could choose both $1/9$ and $2/3$ as relevant objects. Considering from P's perspective $\delta = 0.2$, the CDM for $1/9$ will correspond to $\overset{\diamond}{\uparrow}$ because $\mu_{A@P}(1/9) - \mu_{A@Q}(1/9) = 0.67$ is greater than δ. Likewise, the CDM for $2/3$ will correspond to $\overset{\diamond}{\uparrow}$ because $\mu_{A@P}(2/3) - \mu_{A@Q}(2/3) = 1$ is greater than δ. Thus, a CDP for AAFs $\mu_{A@P}(x_i)$ and $\mu_{A@Q}(x_i)$ from P's perspective will be $\overset{\diamond}{\uparrow}\overset{\diamond}{\uparrow}$.

After a CDP have been obtained, one could assign a weight to it according to his/her strategy to build it. Such a weight will be considered as an approximate value for a CAF. For instance, $\overset{\diamond}{\uparrow}\overset{\diamond}{\uparrow}$ denotes a big difference, thus, P could assign the weight 0 to it. Hence, a CAF for AAFs $\mu_{A@P}$ and $\mu_{A@Q}$ from P's perspective will be $\Delta_{\mu_A:P,Q@P} = \text{weight}(\overset{\diamond}{\uparrow}\overset{\diamond}{\uparrow}) = 0$.

At this point, using an approximate CAF one could compare his/her AAF with a *regular* appraisal function given by someone else. To do so, we rewrite (6.11) as

$$\mathrm{sim}_{\ell@P}(\hat{A}_{@P}, A_{@Q}) = 1 - \frac{1}{n}\sum_{i=1}^{n} |\ell_{\mathcal{C}@P}(x_i) - \ell_{\mathcal{C}@Q}(x_i) \cdot \Delta_{\mathcal{C}:P,Q@P}|, \tag{6.12}$$

where $A_{@Q} = \{\langle x_i, \ell_{\mathcal{C}@Q}(x_i)\rangle | \, (x_i \in X) \wedge (\ell_{\mathcal{C}@Q}(x_i) \in I)\}$ is a set representing a regular appraisal function $\ell_{\mathcal{C}@Q}(x_i)$.

6.4.3 Augmented appraisal degrees in (Atanassov) intuitionistic fuzzy sets

There are cases in which a person could independently judge an object fulfilling a membership and a non-membership criteria at the same time. For example, one could judge an object x fulfilling a *membership* criterion related to a concept A if he/she detects some features in x that are similar to the features in his/her prototype for A; simultaneously, he/she could judge x fulfilling a *non-membership* criterion if he/she detects some features in x that does not belong to his/her prototype for A. This particular situation was considered by Atanassov in the definition of the IFS concept (see Section 6.2.2.1).

As seen throughout this chapter, when two (or more) persons judge an object x satisfying a membership (or non-membership) criterion related to a concept A, they could focus on distinct (collections of) features according to their individual prototypes for A. Hence, to manage different points of view within an IFS, we consider including AADs into its definition as follows:

Definition 6.8 (Augmented (Atanassov) Intuitionistic Fuzzy Set) *Consider a collection of objects $X = \{x_1, \cdots, x_n\}$, where each $x_i \in X$ has a collection of features $\mathcal{F}_i$. Also consider a concept A and a person P. Assume $I = [0,1]$ and $\mathcal{F} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_n$. Let $\hat{\mu}_{A@P}(x_i) = \langle \mu_{A@P}(x_i), F_{\mu_A@P}(x_i)\rangle$ and $\hat{\nu}_{A@P}(x_i) = \langle \nu_{A@P}(x_i), F_{\nu_A@P}(x_i)\rangle$ in $\langle I, \mathcal{F}\rangle$ be two AADs of x_i satisfying, respectively, a* membership *criterion such as "be compatible with the definition of A" and a* non-membership *criterion such as "be incompatible with the definition of A" according to the standpoint of P. A collection $\hat{A}_{@P}$ that denotes the correspondence between each $x_i \in X$ and both $\hat{\mu}_{A@P}(x_i)$ and $\hat{\nu}_{A@P}(x_i)$ such that*

$$\begin{aligned}\hat{A}_{@P} = \{ & \langle x_i, \hat{\mu}_{A@P}(x_i), \hat{\nu}_{A@P}(x_i)\rangle \mid (x_i \in X) \\ & \wedge (0 \le \mu_{A@P}(x_i) + \nu_{A@P}(x_i) \le 1)\}, \end{aligned} \tag{6.13}$$

is called an augmented (Atanassov) intuitionistic fuzzy set*, AAIFS for short.*

To show how an AAIFS could manage appraisals from different perspectives, let us study the following example. Consider a collection $X = \{x_1\}$ and two AAIFSs:

$$\begin{aligned}\hat{A}_{@P} &= \{\langle x_1, \langle 0.625, F_{\mu_A@P}\rangle, \langle 0.375, F_{\nu_A@P}\rangle\rangle\} \text{ and} \\ \hat{A}_{@Q} &= \{\langle x_1, \langle 0.125, F_{\mu_A@Q}\rangle, \langle 0.500, F_{\nu_A@Q}\rangle\rangle\},\end{aligned}$$

given by persons P and Q respectively. Figure 6.4 depicts the geometrical interpretations based on *IFS-interpretational triangles* [7] of the AAIFS elements:

$$\begin{aligned}\bigcirc &= \langle x_1, \langle 0.625, F_{\mu_A@P}\rangle, \langle 0.375, F_{\nu_A@P}\rangle\rangle \text{ and} \\ \square &= \langle x_1, \langle 0.125, F_{\mu_A@Q}\rangle, \langle 0.500, F_{\nu_A@Q}\rangle\rangle,\end{aligned}$$

i.e., $\bigcirc \in \hat{A}_{@P}$ and $\square \in \hat{A}_{@Q}$. The following cases are shown:

- In Figure 6.4a, the features considered for appraising the membership criterion, as well as the features for appraising the non-membership criterion are the same in both perspectives, i.e., $F_{\mu_A@P} = F_{\mu_A@Q}$ and $F_{\nu_A@P} = F_{\nu_A@Q}$. Thus, $\mu_{A@Q}(x_1) = 0.125$ is seen as it is from the perspective of P, i.e., $\lfloor \mu_{A@Q}(x_1)\rfloor_{@P} = \mu_{A@Q}(x_1)$ since $\Delta_{\mu_A:P,Q@P} = 1$. Likewise, $\nu_{A@Q}(x_1) = 0.500$ is seen as it is from the perspective of P, i.e., $\lfloor \nu_{A@Q}(x_1)\rfloor_{@P} = \nu_{A@Q}(x_1)$ since $\Delta_{\nu_A:P,Q@P} = 1$. We could say that this situation corresponds to the original definition of an IFS.
- In Figure 6.4b, only the features considered for appraising the non-membership criterion are the same in both perspectives, i.e., $F_{\nu_A@P} = F_{\nu_A@Q}$. Since $\Delta_{\nu_A:P,Q@P} = 1$ and $\Delta_{\mu_A:P,Q@P} \neq 1$, only $\lfloor \nu_{A@Q}(x_1)\rfloor_{@P} = \nu_{A@Q}(x_1)$ holds.
- In Figure 6.4c, only the features considered for appraising the membership criterion are the same in both perspectives, i.e., $F_{\mu_A@P} = F_{\mu_A@Q}$. Thus, only $\lfloor \mu_{A@Q}(x_1)\rfloor_{@P} = \mu_{A@Q}(x_1)$ holds.
- In Figure 6.4d, there are one or more features used to appraise the membership criterion from P that are different from those used by Q. The

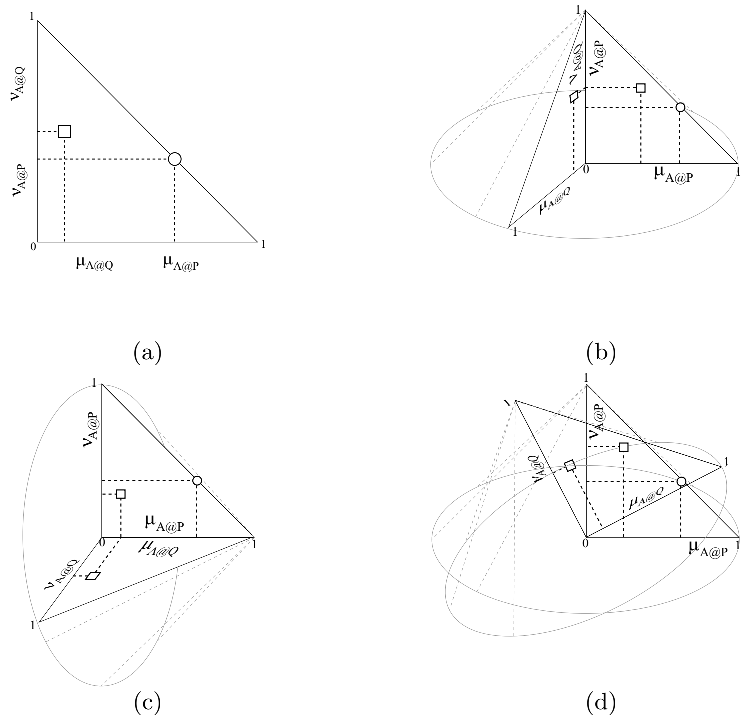


Figure 6.4: Geometrical interpretation based on the IFS-interpretational triangles of two AAIFS elements, $\bigcirc \in \hat{A}_{@P}$ and $\square \in \hat{A}_{@Q}$, given from the perspectives of P and Q respectively, where $\bigcirc$ corresponds to $\langle x_1, \langle 0.625, F_{\mu_A @P}\rangle, \langle 0.375, F_{\nu_A @P}\rangle\rangle$, and $\square$ corresponds to $\langle x_1, \langle 0.125, F_{\mu_A @Q}\rangle, \langle 0.500, F_{\nu_A @Q}\rangle\rangle$.

same happens with the features used to appraise the non-membership criterion. This means that $F_{\mu_A @P} \neq F_{\mu_A @Q}$ and $F_{\nu_A @P} \neq F_{\nu_A @Q}$. Accordingly, $\lfloor \mu_{A@Q}(x_1) \rfloor_{@P} \neq \mu_{A@Q}(x_1)$ and $\lfloor \nu_{A@Q}(x_1) \rfloor_{@P} \neq \nu_{A@Q}(x_1)$ hold since $\Delta_{\mu_A : P, Q@P} \neq 1$ and $\Delta_{\nu_A : P, Q@P} \neq 1$.

6.4.3.1 Comparing augmented (Atanassov) intuitionistic fuzzy sets

The approach proposed in Section 6.4.2.1 to compare two AAFs is also suited for comparing two AAIFSs. Hence, to obtain an $\ell-$similarity measure of the similarity between two AAIFSs, say $\hat{A}_{@P}$ and $\hat{A}_{@Q}$, from a particular perspective, say the perspective of person P, we rewrite (6.11) like this:

$$\mathrm{sim}_{\ell @P}(\hat{A}_{@P}, \hat{A}_{@Q}) = 1 - \frac{1}{n}\sum_{i=1}^{n} |\mathrm{dif}_{\ell @P}\left(\mathbf{p_i}, \mathbf{q_i}\right)|, \tag{6.14}$$

where:

$$\mathbf{p_i} = \begin{pmatrix} \mu_{A@P}(x_i) & + & \alpha_{A@P} \cdot h_{A@P}(x_i) \\ \nu_{A@P}(x_i) & + & (1-\alpha_{A@P}) \cdot h_{A@P}(x_i) \end{pmatrix} \text{and}$$

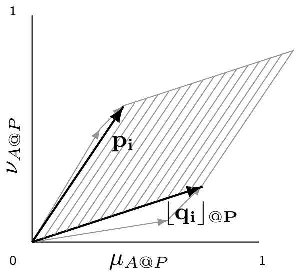


Figure 6.5: Idea behind $\mathrm{dif}_{\ell@P}\left(\mathbf{p_i}, \mathbf{q_i}\right)$: the larger the area of the parallelogram formed by vectors $\mathbf{p_i}$ and $\lfloor \mathbf{q_i} \rfloor_{@P}$, the larger the difference between $\mathbf{p_i}$ and $\mathbf{q_i}$ as seen from P.

$$\mathbf{q_i} = \begin{pmatrix} \mu_{A@Q}(x_i) & + & \alpha_{A@Q} \cdot h_{A@Q}(x_i) \\ \nu_{A@Q}(x_i) & + & (1 - \alpha_{A@Q}) \cdot h_{A@Q}(x_i) \end{pmatrix}$$

are *vector interpretations* (see Section 3.3.1) of $\langle x_i, \hat{\mu}_{A@P}(x_i), \hat{\nu}_{A@P}(x_i) \rangle \in \hat{A}_{@P}$ and $\langle x_i,\, \hat{\mu}_{A@Q}(x_i),\, \hat{\nu}_{A@Q}(x_i) \rangle \in \hat{A}_{@Q}$ in that order;

$$h_{A@P}(x_i) = 1 - \mu_{A@P}(x_i) - \nu_{A@P}(x_i) \text{ and}$$

$$h_{A@Q}(x_i) = 1 - \mu_{A@Q}(x_i) - \nu_{A@Q}(x_i)$$

are the hesitation margins from the perspectives of P and Q correspondingly; $\alpha_{A@P}, \alpha_{A@Q} \in [0, 1]$ are *hesitation splitters* that split any hesitation about the membership and non-membership for x_i from P and Q respectively (see Section 3.3.2). So,

$$\begin{aligned} \mathrm{dif}_{\ell@P}\left(\mathbf{p_i}, \mathbf{q_i}\right) = {} & \left(\mu_{A@P}\left(x_i\right) - \lfloor \mu_{A@Q}\left(x_i\right) \rfloor_{@P}\right) \\ & + \left(\alpha_{A@P} \cdot h_{A@P}(x_i) - \alpha_{A@Q} \cdot \lfloor h_{A@Q}(x_i) \rfloor_{@P}\right), \end{aligned} \tag{6.15}$$

is the *spot difference* between $\mathbf{p_i}$ and $\mathbf{q_i}$ as seen from the perspective of P (cf. (3.8)); and

$$\lfloor h_{A@Q}(x_i) \rfloor_{@P} = 1 - \lfloor \mu_{A@Q}(x_i) \rfloor_{@P} - \lfloor \nu_{A@Q}(x_i) \rfloor_{@P} \tag{6.16}$$

is the hesitation margin resulting from Q as seen from the standpoint of P.

The idea behind (6.15) is depicted in Figure 6.5. Vector $\lfloor \mathbf{q_i} \rfloor_{@P}$ corresponds to vector $\mathbf{q_i}$ as seen from perspective P, i.e.,

$$\lfloor \mathbf{q_i} \rfloor_{@P} = \begin{pmatrix} \lfloor \mu_{A@Q}(x_i) \rfloor_{@P} & + & \alpha_{A@Q} \cdot \lfloor h_{A@Q}(x_i) \rfloor_{@P} \\ \lfloor \nu_{A@Q}(x_i) \rfloor_{@P} & + & (1 - \alpha_{A@Q}) \cdot \lfloor h_{A@Q}(x_i) \rfloor_{@P} \end{pmatrix}. \tag{6.17}$$

The area of the parallelogram formed by vectors $\mathbf{p_i}$ and $\lfloor \mathbf{q_i} \rfloor_{@P}$ is used as a reference to measure the difference between them: the larger the area of the parallelogram, the larger the difference between $\mathbf{p_i}$ and $\mathbf{q_i}$ as seen from the perspective of P. Thereby, the largest area is determined by vectors $\mathbf{m_f} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{n_f} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Therefore, (6.15) could be seen as a consequence of

$$\mathrm{dif}_{\ell @ P}\left(\mathbf{p_i}, \mathbf{q_i}\right) = \frac{\mathbf{p_i} \times \lfloor \mathbf{q_i} \rfloor_{@P}}{\mathbf{m_f} \times \mathbf{n_f}}, \tag{6.18}$$

where $\times$ denotes the vector product.

Semantically, (6.15) could be interpreted in a similar way as (3.8): while the former part of the expression denotes that the difference between $\mathbf{p_i}$ and $\mathbf{q_i}$ is determined by the appraisals of the membership criterion as seen from the perspective of P, the latter part denotes that the difference is also influenced by any doubt about the membership and the non-membership criteria. Moreover, the latter part could be affected by managing both hesitation splitters $\alpha_{A@P}$ and $\alpha_{A@Q}$. For instance, a difference is not affected by doubts if $\alpha_{A@P} = 0$ and $\alpha_{A@Q} = 0$ (see Figure 6.6a); on the other hand, a difference is fully affected by any doubt if $\alpha_{A@P} = 1$ and $\alpha_{A@Q} = 1$ (see Figure 6.6b).

As in (3.10), one could apply the same rule for both splitters, i.e., $\alpha_{A@P} = \alpha_{A@Q} = \alpha$. In such a case (6.15) can then be expressed by

$$\begin{aligned} \mathrm{dif}^{\alpha}_{\ell @ P}\left(\mathbf{p_i}, \mathbf{q_i}\right) = &\left(\mu_{A@P}\left(x_i\right) - \lfloor \mu_{A@Q}\left(x_i\right) \rfloor_{@P}\right) \\ &+ \alpha \cdot \left(h_{A@P}(x_i) - \lfloor h_{A@Q}(x_i) \rfloor_{@P}\right). \end{aligned} \tag{6.19}$$

In a similar way, one can use the membership and non-membership hesitation splitters introduced in (3.11) to define

$$\begin{aligned} dif^{\alpha,\beta}_{\ell @ P}(\mathbf{p_i}, \mathbf{q_i}) = &\left(\mu_{A@P}(x) + \alpha \cdot h_{A@P}(x)\right) \cdot \left(\lfloor \nu_{A@Q}(x) \rfloor_{@P} + \beta \cdot \lfloor h_{A@Q}(x) \rfloor_{@P}\right) \\ &- \left(\lfloor \mu_{A@Q}(x) \rfloor_{@P} + \alpha \cdot \lfloor h_{A@Q}(x) \rfloor_{@P}\right) \cdot \left(\nu_{A@P}(x) + \beta \cdot h_{A@P}(x)\right) \end{aligned} \tag{6.20}$$

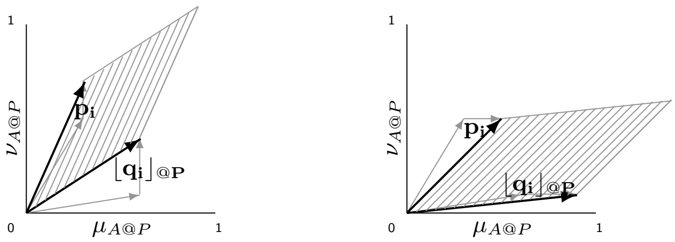
Consequently, (6.14) can be expressed by

$$\mathrm{sim}^{\alpha}_{\ell @ P}(\hat{A}_{@P}, \hat{A}_{@Q}) = 1 - \frac{1}{n} \sum_{i=1}^{n} \left|\mathrm{dif}^{\alpha}_{\ell @ P}\left(\mathbf{p_i}, \mathbf{q_i}\right)\right| \tag{6.21}$$

or

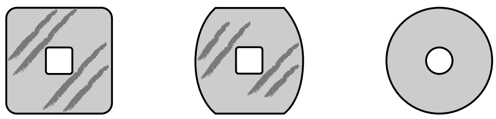
$$\mathrm{sim}^{\alpha,\beta}_{\ell @ P}(\hat{A}_{@P}, \hat{A}_{@Q}) = 1 - \frac{1}{n} \sum_{i=1}^{n} \left|\mathrm{dif}^{\alpha,\beta}_{\ell @ P}\left(\mathbf{p_i}, \mathbf{q_i}\right)\right|. \tag{6.22}$$

An interesting aspect of (6.14), (6.21) and (6.22) is the allowance of the specification of two distinct CAFs, one for the membership and the other for non-membership. Thus, it is possible to take account of the kind of situations depicted in Figure 6.4 during a comparison between two AAIFSs. In the next section, we will show how to do so when comparing two AAIFSs that correspond to two collections of XBEs.



(a) $\alpha = 0$. (b) $\alpha = 1$.

Figure 6.6: Managing α hesitation splitter.



(a) Alice's. (b) Bob's. (c) Chloe's.

Figure 6.7: How a Grandma's cookie looks like according to each relative.

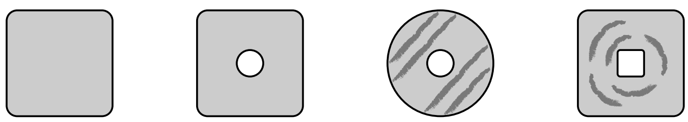
6.4.3.2 Augmented (Atanassov) intuitionistic fuzzy sets and XBEs

As indicated in Section 6.1, the main purpose of this chapter is to handle in a better way the connotative meaning of collections of XBEs in order to perform a meaningful comparison between any two of them. To show how to do that using AAIFSs, let us consider again the *Grandma's cookies* example presented in Section 3.2.1 – to help the reader, the cookies introduced in that example, which were depicted in Figure 3.1 and Figure 3.2, are depicted again in Figure 6.7 and Figure 6.8 respectively.

Notice in the Grandma's cookies example that the mental picture of a Grandma's cookie from each cousin (see Figure 6.7) depends on his/her particular experience with such cookies. Thus, since there is nothing but individual memories used as referents for the evaluations in Table 6.2, we deem them to be XBEs.

To model such XBEs, we use the AAIFS concept as follows:

- The cookies in Figure 6.8 correspond to $X = \{x_1, x_2, x_3, x_4\}$ – e.g., *cookie* 1 will correspond to x_1 and so on.



(a) *cookie* 1. (b) *cookie* 2. (c) *cookie* 3. (d) *cookie* 4.

Figure 6.8: Do these cookies look like a *Grandma's cookie?*.

Table 6.2: To which degree each cookie in Figure 6.8 is seen or not as a Grandma's cookie by each cousin (Grandma's cookie example)?

	yes		*no*	
cookie	level	reason(s)	level	reason(s)
1	0.6	square shape	0.3	no square hole, no linear icing
2	0.7	square shape	0.3	round hole, no linear icing
3	0.2	linear icing	0.8	round shape, round hole
4	0.9	square shape, square hole	0.1	curved icing

(a) Evaluations by Alice.

	yes		*no*	
cookie	level	reason(s)	level	reason(s)
1	0.4	square shape	0.3	no square hole, no linear icing
2	0.5	square shape	0.3	round hole, no linear icing
3	0	linear icing	0.9	round shape, round hole
4	0.7	square shape, square hole	0.2	curved icing

(b) Evaluations by Bob.

	yes		*no*	
cookie	level	reason(s)	level	reason(s)
1	0.6	no icing	0.3	square shape, no (round) hole
2	0.7	no icing, round hole	0.3	square shape
3	0.2	round shape, round hole	0.8	(linear) icing
4	0.1	(square) hole	0.9	(curved) icing

(c) Evaluations by Chloe.

- The features in, e.g., *cookie* 4 correspond to $\mathcal{F}_4 = \{$'*square shape*', '*square hole*', '*curved icing*'$\}$.
- The criterion for the evaluation is "be compatible with the way in which A is perceived," where A represents *a Grandma's cookie.*
- It is considered that the evaluations of Alice, Bob and Chloe (see Table 6.2) respectively correspond to the AAIFSs $\hat{A}_{@P}$, $\hat{A}_{@Q}$ and $\hat{A}_{@R}$. For instance, using the data in Table 6.2a, a complete view of $\hat{A}_{@P}$ (Alice's perspective) is

$$\begin{aligned}\hat{A}_{@P} = \{ \; &\langle x_1, \langle 0.6, F_{\mu_A @P}(x_1)\rangle , \langle 0.3, F_{\nu_A @P}(x_1)\rangle\rangle \, , \\ &\langle x_2, \langle 0.7, F_{\mu_A @P}(x_2)\rangle , \langle 0.3, F_{\nu_A @P}(x_2)\rangle\rangle \, , \\ &\langle x_3, \langle 0.2, F_{\mu_A @P}(x_3)\rangle , \langle 0.8, F_{\nu_A @P}(x_3)\rangle\rangle \, , \\ &\langle x_4, \langle 0.9, F_{\mu_A @P}(x_4)\rangle , \langle 0.1, F_{\nu_A @P}(x_4)\rangle\rangle \} \, ,\end{aligned}$$

  where $F_{\mu_A @P}(x_i)$ and $F_{\nu_A @P}(x_i)$ correspond to the collections of 'yes'-reason(s) and 'no'-reason(s) given for cookie x_i respectively.

As was mentioned throughout the chapter, a comparison between two collections of XBEs depends on the perspective taken as a referent, thus, we should adopt a particular point of view to compare the evaluations from Alice, Bob and Chloe. Let us adopt Alice's perspective. To compare her evaluations with

Bob's evaluations using (6.21), i.e., $\text{sim}^{\alpha}_{\ell@P}(\hat{A}_{@P}, \hat{A}_{@Q})$, first it is necessary to establish a CAF for each criterion: $\Delta_{\mu_A:P,Q@P}$ for the membership criterion and $\Delta_{\nu_A:P,Q@P}$ for the non-membership criterion. To do so, we compare the reasons in each evaluation given by Alice and Bob for each criterion, i.e., $F_{\mu_A@P}(x_i)$ vs. $F_{\mu_A@Q}(x_i)$ and $F_{\nu_A@P}(x_i)$ vs. $F_{\nu_A@Q}(x_i)$ for each $x_i \in X$. Since the reasons are the same (see Tables 6.2a and 6.2b), we assign $\Delta_{\mu_A:P,Q@P} = 1$ and $\Delta_{\nu_A:P,Q@P} = 1$ (cf. the situation depicted in Figure 6.4a) – here, we highlight that it is also possible to determine the values of both CAFs using an approximation method as the one given in Section 6.4.2.2. Now, using vector interpretations of each evaluation in $\hat{A}_{@P}$ and $\hat{A}_{@Q}$, we compute the spot difference for each $x_i \in X$, i.e., $\text{dif}_{\ell@P}(\mathbf{p_i}, \mathbf{q_i})$. For instance, the vector interpretations from Alice and Bob's evaluations for x_1 are $\mathbf{p_1} = \begin{pmatrix} 0.6 & + & \alpha \cdot 0.1 \\ 0.3 & + & (1-\alpha)\cdot 0.1 \end{pmatrix}$ and $\mathbf{q_1} = \begin{pmatrix} 0.4 & + & \alpha \cdot 0.3 \\ 0.3 & + & (1-\alpha)\cdot 0.3 \end{pmatrix}$ respectively. With (6.17), the latter corresponds to $\lfloor \mathbf{q_1} \rfloor_{@P} = \begin{pmatrix} \lfloor 0.4 \rfloor_{@P} & + & \alpha \cdot \lfloor 0.3 \rfloor_{@P} \\ \lfloor 0.3 \rfloor_{@P} & + & (1-\alpha)\cdot \lfloor 0.3 \rfloor_{@P} \end{pmatrix}$ from Alice's perspective. Using (6.6) and (6.16) with $\Delta_{\mu_A:P,Q@P}$ and $\Delta_{\nu_A:P,Q@P}$, we obtain $\lfloor \mathbf{q_1} \rfloor_{@P} = \begin{pmatrix} 0.4 \cdot 1 & + & \alpha \cdot (1 - 0.4 \cdot 1 - 0.3 \cdot 1) \\ 0.3 \cdot 1 & + & (1-\alpha)\cdot (1 - 0.4 \cdot 1 - 0.3 \cdot 1) \end{pmatrix}$ and, with (6.19), we obtain $\text{dif}_{\ell@P}(\mathbf{p_1}, \mathbf{q_1}) = (0.6 - 0.4) + \alpha \cdot (0.1 - 0.3) = 0.2 - 0.2\alpha$. Likewise, $\text{dif}_{\ell@P}(\mathbf{p_2}, \mathbf{q_2}) = 0.2$, $\text{dif}_{\ell@P}(\mathbf{p_3}, \mathbf{q_3}) = 0.2 - 0.1\alpha$ and $\text{dif}_{\ell@P}(\mathbf{p_4}, \mathbf{q_4}) = 0.2 - 0.1\alpha$ are obtained. Finally, using (6.21), we obtain $\text{sim}^{\alpha}_{\ell@P}(\hat{A}_{@P}, \hat{A}_{@Q}) = 0.8 + 0.1\alpha$ as a similarity value for Alice vs. Bob's evaluations as seen from Alice's perspective.

In a similar way, to compare from Alice's perspective her evaluations with Chloe's, it is necessary to establish the values for $\Delta_{\mu_A:P,R@P}$ and $\Delta_{\nu_A:P,R@P}$. To do so, the approximation method given in Section 6.4.2.2 is used. Choosing x_4 and x_3 as representative objects for membership and non-membership respectively, with $\delta = 0.2$ the CDM for the membership criterion in x_4 is ⫯ because $\mu_{A@P}(x_4) - \mu_{A@R}(x_4) = 0.9 - 0.1 = 0.8$, and the CDM for the non-membership criterion in x_3 is ⌽ because $\nu_{A@P}(x_3) - \nu_{A@R}(x_3) = 0.8 - 0.8 = 0$. With the corresponding CDMs, we establish ⫯ as the CDP for $\mu_{A@P}(x_i)$ and $\mu_{A@R}(x_i)$, and ⌽ for $\nu_{A@P}(x_i)$ and $\nu_{A@R}(x_i)$. Accordingly, we set $\Delta_{\mu_A:P,R@P} = \text{weight}(⫯) = 0$ and $\Delta_{\nu_A:P,R@P} = \text{weight}(⌽) = 1$ (cf. the situation depicted in Figure 6.4b). Following the above procedure, $\text{dif}_{\ell@P}(\mathbf{p_1}, \mathbf{r_1}) = 0.6$, $\text{dif}_{\ell@P}(\mathbf{p_2}, \mathbf{r_2}) = 0.7$, $\text{dif}_{\ell@P}(\mathbf{p_3}, \mathbf{r_3}) = 0.2$ and $\text{dif}_{\ell@P}(\mathbf{p_4}, \mathbf{r_4}) = 0.9$ are obtained. Finally, we obtain $\text{sim}^{\alpha}_{\ell@P}(\hat{A}_{@P}, \hat{A}_{@R}) = 0.4$ as the similarity value for Alice vs. Chloe's evaluations as seen from Alice's perspective. This means that, even with a no-doubts comparison strategy, i.e., a strategy in which $\alpha = 0$, the evaluations of Alice-vs.-Bob are more similar than the evaluations of Alice-vs.-Chloe.

As might be observed from the example, modeling collections of XBEs using AAIFSs allows for a semantic richer and, hence, a more reliable comparison between any two of them. Moreover, it could be noticed that it is possible to compare an AAIFS with an IFS, which is useful for scenarios where only one evaluator provides hints about his/her judgments. In the next section we will present some potential applications in which it is beneficial to use such semantic richer comparisons.

6.4.4 Potential applications of augmented appraisal degrees

So far, we have studied how an AAD can characterize not only the extent but also the context of an XBE. Moreover, we have seen that, although XBEs could be fairly subjective, such a characterization can make a comparison between any two of them a more reliable act. Therefore, AADs can potentially be applied in those situations where a comparison of subjective answers to an (implicit or explicit) evaluation request is needed.

As was mentioned in the introduction to this chapter, a citizen science project is such a situation. Since a large number of answers provided by people with different background are expected in this kind of (crowdsourcing) projects [27], the proposed augmented framework can be used to filter, classify or arrange the answers. Moreover, it could be used in a validation process in which the quality of the provided data depends on a particular understanding of the concept under consideration – e.g., [28, 29] are some works related to data quality control to which the augmented framework could be applied.

Processing huge amounts of subjective data is another situation in which the augmented framework can be used. As will be shown in Chapter 7, the augmented framework can be used to obtain an approximation of the level to which the contexts of subjective (fuzzy) judgments on social media content are perceived as alike.

As will be seen in Chapter 9, the augmented framework could also be applied is group decision making problems involving heterogeneous experts (e.g., [30]). In this case, AADs could be used to model the experts' preferences in such a way that, using the recorded hints, a moderator could know the reasons behind each preference and, thus, conduct a more informed consensus process.

The construction of fuzzy ontologies based on information collected from different sources in social media (e.g., [31, 32]) is another potential application of AADs. In this case, the proposed augmented framework could help to build a kind of *contextual* fuzzy ontology that allows the users to perform queries according to their particular understandings of the constituent concepts.

It is worth mentioning that, since the proposed framework aims to deal in a better way with comparisons of XBEs, its applicability can be assessed according to, e.g., the level to which a comparison between AADs reflects what is perceived. For instance, a test analogous to the one presented in Chapter 4, in which several similarity measures for IFSs were tested in comparisons of (simulated) XBEs, could be used to assess how well the computed similarity between two AAIFSs reflects the perceived similarity between them.

6.5 Conclusions

In this chapter, we studied a novel generalization of an appraisal degree, e.g., a membership (or non-membership) grade, to denote in a better way the connotative meaning in an *experience-based evaluation* (XBE), i.e., an evaluation resulting from what one has learned or understood about a particular topic by experience (Research Question $Q1$). If a membership (or non-membership) grade denotes to *which degree* a membership (or non-membership) criterion is fulfilled by an object x, the generalization additionally hints *why* such a criterion is fulfilled by x. Here, x represents a person, a notion, or something that exists by itself; and a membership (or non-membership) criterion represents a reason for making a judgment about the membership (or non-membership) of x. From a psychological point of view, a person could focus on particular inherent characteristics, i.e., features, of an object x in order to make a judgment about it. Such features are deemed as *hints of a judgment* in the generalization. Due to this, the generalization is called an *augmented appraisal degree*, AAD for short.

An AAD allows for dealing with a comparison between two XBEs, which could be affected not just by the magnitude of each appraisal, but also by the features that are focused on according to individual understandings of a concept. Its advantages can be briefed as follows:

- An AAD takes account of a human behavior in which, together with an appraisal level, hints of the appraisal are given.
- An AAD can be seen from different perspectives through an *'as seen from'* operator (Research Question $Q3$).
- AADs can be compared with each other from a particular perspective using *numerical comparison operators* (e.g., $=$, $>$ or $<$) or their *fuzzy* counterparts (e.g., *'approximately equal'*, *'not much larger than'*, or *'much smaller than'*) (Research Questions $Q3$ and $Q4$).
- A collection of AADs representing judgments of a collection X from a particular point of view can be denoted by an *augmented appraisal function*, AAF for short (Research Question $Q1$).
- Through AAFs, collections of XBEs given from different perspectives can be mathematically represented (Research Question $Q1$).
- By means of AAFs, an (Atanassov) intuitionistic fuzzy set can be extended to cope with different points of view while handling XBEs (Research Questions $Q2$, $Q3$ and $Q4$).
- By means of AAFs, a similarity comparison between two collections of XBEs becomes a more reliable act (Research Question $Q2$).

Important aspects that will be studied in the next two chapters are (i) "*how to handle* flat *XBEs* (i.e., how to handle XBEs in which the hints of a judgment have not been recorded) and (ii) "*how to handle* unrequested *XBEs* (i.e., how to handle XBEs when there is no explicit evaluation request).

References

[1] Eric Hand. *Volunteer army catches interstellar dust grains.* Nature, 466(August):685–687, March 2010.

[2] Michael F. Goodchild. *Citizens as sensors: the world of volunteered geography.* GeoJournal, 69(4):211–221, 2007.

[3] Christopher S. Lowry and Michael N. Fienen. *CrowdHydrology: crowdsourcing hydrologic data and engaging citizen scientists.* Ground water, 51(1):151–6, 2013.

[4] Mark H. Palmer and Scott Kraushaar. *Volunteered Geographic Information, Actor-Network Theory, and Severe Storm Reports.* In Daniel Sui, Sarah Elwood, and Michael Goodchild, editors, Crowdsourcing Geographic Knowledge. Springer Netherlands, 2013.

[5] Benjamin M. Good and Andrew I. Su. *Crowdsourcing for bioinformatics.* Bioinformatics (Oxford, England), 29(16):1925–33, August 2013.

[6] Krassimir T. Atanassov. *Intuitionistic fuzzy sets.* Fuzzy sets and Systems, 20(1):87–96, 1986.

[7] Krassimir T. Atanassov. *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[8] Eulalia Szmidt and Janusz Kacprzyk. *Distances between intuitionistic fuzzy sets.* Fuzzy Sets and Systems, 114(3):505–518, September 2000.

[9] Eulalia Szmidt and Janusz Kacprzyk. *A concept of similarity for intuitionistic fuzzy sets and its use in group decision making.* In Fuzzy Systems, 2004. Proceedings. 2004 IEEE International Conference on, volume 2, pages 1129–1134. IEEE, 2004.

[10] Eulalia Szmidt and Janusz Kacprzyk. *A New Concept of a Similarity Measure for Intuitionistic Fuzzy Sets and Its Use in Group Decision Making.* In Modeling Decisions for Artificial Intelligence, volume 3558 of *Lecture Notes in Computer Science*, pages 272–282. Springer Berlin Heidelberg, 2005.

[11] Ioannis K. Vlachos and George D. Sergiadis. *Intuitionistic fuzzy information - Applications to pattern recognition.* Pattern Recognition Letters, 28(2):197 – 206, 2007.

[12] Eulalia Szmidt and Marta Kukier. *Atanassov Intuitionistic Fuzzy Sets in Classification of Imbalanced and Overlapping Classes.* In Panagiotis Chountas, Ilias Petrounias, and Janusz Kacprzyk, editors, Intelligent Techniques and Tools for Novel System Architectures, volume 109 of *Studies in Computational Intelligence*, pages 455–471. Springer Berlin Heidelberg, 2008.

[13] Marcelo Loor and Guy De Tré. *Vector based similarity measure for intuitionistic fuzzy sets.* In Krassimir T. Atanassov, Michał Baczyński, Józef Drewniak, Janusz Kacprzyk, Maciej Krawczak, Eulalia Szmidt, Maciej Wygralak, and Sławomir Zadrożny, editors, Modern approaches in fuzzy sets, intuitionistic fuzzy sets, generalized nets and related topics : volume I : foundations, pages 105–127. SRI-PAS, 2014.

[14] Eulalia Szmidt and Janusz Kacprzyk. *Geometric similarity measures for the intuitionistic fuzzy sets.* In 8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13), pages 840–847. Atlantis Press, 2013.

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

[16] Ronald R. Yager and Ali M. Abbasov. *Pythagorean Membership Grades, Complex Numbers, and Decision Making.* International Journal of Intelligent Systems, 28(5):436–452, 2013.

[17] Ronald R. Yager. *Pythagorean fuzzy subsets.* In IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint, pages 57–61, June 2013.

[18] Tom Matthé and Guy De Tré. *Bipolar query satisfaction using satisfaction and dissatisfaction degrees: bipolar satisfaction degrees.* In Proceedings of the 2009 ACM symposium on Applied Computing, pages 1699–1703. ACM, 2009.

[19] Guy De Tré, Sławomir Zadrożny, Tom Matthé, Janusz Kacprzyk, and Antoon Bronselaer. *Dealing with Positive and Negative Query Criteria in Fuzzy Database Querying.* In Troels Andreasen, Ronald R. Yager, Henrik Bulskov, Henning Christiansen, and Henrik L. Larsen, editors, Flexible Query Answering Systems, volume 5822 of *Lecture Notes in Computer Science*, pages 593–604. Springer Berlin Heidelberg, 2009.

[20] Didier Dubois, Walenty Ostasiewicz, and Henri Prade. *Fuzzy Sets: History and Basic Notions.* In Didier Dubois and Henri Prade, editors, Fundamentals of Fuzzy Sets, volume 7 of *The Handbooks of Fuzzy Sets Series*, pages 21–124. Springer US, 2000.

[21] Amos Tversky. *Features of similarity.* Psychological review, 84(4):327, 1977.

[22] Lotfi A. Zadeh. *Toward a restriction-centered theory of truth and meaning (RCT).* Information Sciences, 248(0):1 – 14, 2013.

[23] Lotfi A. Zadeh. *Is there a need for fuzzy logic?* Information Sciences, 178(13):2751 – 2779, 2008.

[24] Marcelo Loor and Guy De Tré. *Connotation-differential Prints - Comparing What Is Connoted Through (Fuzzy) Evaluations.* In Proceedings of the International Conference on Fuzzy Computation Theory and Applications, pages 127–136, 2014.

[25] Lotfi A. Zadeh. *A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges.* Journal of Cybernetics, 2(3):4–34, 1972.

[26] Sergei Ovchinnikov. *An Introduction to Fuzzy Relations.* In Didier Dubois and Henri Prade, editors, Fundamentals of Fuzzy Sets, volume 7 of *The Handbooks of Fuzzy Sets Series*, pages 233–259. Springer US, 2000.

[27] Daren C. Brabham. *Crowdsourcing.* Mit Press, 2013.

[28] Alycia W. Crall, Gregory J. Newman, Thomas J. Stohlgren, Kirstin A. Holfelder, Jim Graham, and Donald M. Waller. *Assessing citizen science data quality: an invasive species case study.* Conservation Letters, 4(6):433–442, 2011.

[29] Gloria Bordogna, Paola Carrara, Laura Criscuolo, Monica Pepe, and Anna Rampini. *A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science.* Information Sciences, 258:312–327, February 2014.

[30] Ignacio J. Pérez, Francisco J. Cabrerizo, Sergio Alonso, and Enrique Herrera-Viedma. *A new consensus model for group decision making problems with non-homogeneous experts.* Systems, Man, and Cybernetics: Systems, IEEE Transactions on, 44(4):494–498, 2014.

[31] Juan A. Morente-Molinera, Ignacio J. Pérez, Raquel Ureña, and Enrique Herrera-Viedma. *Building and managing fuzzy ontologies with heterogeneous linguistic information.* Knowledge-Based Systems, 88:154 – 164, 2015.

[32] Juan A. Morente-Molinera, Ignacio J. Pérez, Raquel Ureña, and Enrique Herrera-Viedma. *Creating knowledge databases for storing and sharing people knowledge automatically using group decision making and fuzzy ontologies.* Information Sciences, 328:418 – 434, 2016.

Chapter 7

Handling Plain Experience-Based Evaluations

Abstract

In Chapter 6, we described a method that relies on the hints recorded in experience-based evaluations (XBEs) to approximate the level to which the contexts of such XBEs are perceived as alike. By contrast, in this chapter we propose a novel method that uses only the appraisal levels included in XBEs of a specific number of relevant objects to compute such an approximation – here, by '*relevant object*' we mean an object that is a good example of compatibility (or incompatibility) with the concept under analysis. In this regard, the proposed method constitutes a tool to identify, measure and handle context in *plain XBEs*, i.e., XBEs in which only appraisal levels are available. We demonstrate that simulated judgments support the effectiveness of the proposed method.

This chapter is an adapted version of the following manuscript:

- Marcelo Loor and Guy De Tré. *Identifying and Properly Handling Context in Crowdsourcing*. Submitted for publication in *Applied Soft Computing*.

7.1 Introduction

Nowadays companies and organizations such as airlines or governmental agencies are taking decisions about their products or services on the basis of direct feedback or perceived sentiment extracted from social media content. For instance, to make governmental policies more reachable and effective, some governmental agencies are starting to assimilate the messages, videos or images published online about that policies [1, 2]. Yet, a challenging task related to such practices is the assessment of the truthfulness and suitability of used social media content, which may include highly subjective and diverse forms of expressions.

An available option to complete that task is to make use of *crowdsourcing services* in which workers (or contributors) perform such assessments on behalf of a *requester* like an organization, editor or social media professional [3, 4, 5]. However, such crowdsourced assessments can also be highly dependent on the individual experience or knowledge of each evaluator and, therefore, they possibly do not reflect the perception of a requester. For instance, while a particular social media post on the *XYZ Act* could be deemed to be encouraging by an anonymous worker due to its text, a social media professional in a governmental agency may consider that post inappropriate because of its form (see Figure 7.1). In this regard, a practical motivation for our work is to address the question: *how to make the assessment task available only to evaluators with whom a requester shares a similar understanding of the topic under analysis?* (also reconsider Figure 1.5a)

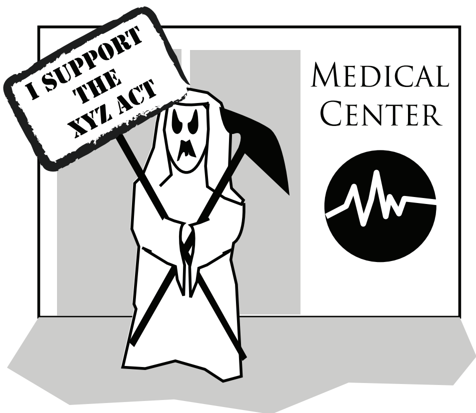


Figure 7.1: Does this post support the XYZ Act?

Although approaches oriented to detect workers who demonstrate abilities to perform a particular task exist, in those approaches usually it is assumed that the assessments (or answers) are unique, independent and unambiguous [6, 7]. Notwithstanding, this assumption can introduce strict restrictions when evaluators are trying to express subjective assessments making it hard for them to formulate proper answers.

To make it easier for an evaluator to make a subjective evaluation, we consider that such an evaluation can be expressed in a way that is not specific and subject to hesitation. For example, one would like to express that the post depicted in Figure 7.1 rather supports the XYZ Act due to the text on it but, at the same time, express that this post denotes a bitter opposition to the XYZ Act because of the drawing on it. In this case, to detect workers sharing a similar understanding with a requester, we suggest to perform *comparisons in which not only the appraisal levels, but also the context of those evaluations is taken into account* – here, by '*context of an evaluation*' we mean the conditions that arise when the evaluation is carried out, which mainly depend on the experience of an evaluator about the concept under analysis.

In Chapter 6 we described an augmented framework for performing such comparisons. In that framework an experience-based evaluation (XBE) is characterized by an *augmented appraisal degree* (AAD), which allows an evaluator to record not only the level, but also some hints on the reasons for his/her appraisal. So, by comparing the hints recorded in two evaluations, one can estimate how similar the contexts of those evaluations are. Nevertheless, recording the reasons of an assessment might be considered burdensome by evaluators in crowdsourcing. Hence, *comparing the contexts of XBEs in which only the appraisal levels have been recorded* still constitutes a specific challenge in this topic.

In this chapter we propose a novel method to address that specific challenge. The method, named *k-well-(un)fitted-specimens method* or *kWFS* for short, relies on the appraisal levels corresponding to the evaluations of a specific number of social media posts that a requester considers to be representative for the concept under study. Indeed, the main idea behind this method indicates that, during an evaluation process, one can detect some posts with features that make them well (un)fitted specimens of the concept under study; thus, if a difference between the understandings of that concept exists, it will be reflected through a difference between the appraisal levels of those posts.

An important and interesting aspect of the kWFS method is that it allows a requester to compute an approximation of the level to which the context of his/her evaluations is similar to the context of the evaluations given by somebody else – a requester may consider this approximation to be an important data quality attribute [8]. Therefore, the kWFS method can be applied in situations where a requester needs to determine how good (or bad) the evaluations given by a particular worker are.

To present the kWFS method, the chapter has been structured as follows. The literature related to this work is presented in the next section. Then, the key ideas and definitions for modeling and comparing the contexts of experience-based evaluations proposed in [9] are summarized. After that, our method is described in detail. Then, we test the effectiveness of the method using simulated experience-based evaluations. We end with some suggestions for future research.

7.2 Related work

Studies about how to detect and recruit qualified workers or contributors can be found in crowdsourcing literature. As was mentioned in Section 1.4.3, most of those studies use a *gold standard* collection containing questions with correct answers to rate the reliability of each worker [7, 10]. Such a collection may be given by an expert [11], may result from inferred correct answers [12, 13, 14], or may result from a process that generate correct answers based on known answers [6]. It may also result from agreement among the workers [15].

Those studies usually assume that the answers are precise and unaffected by any difference in understandings that the workers may have. With that assumption, one can compute the reliability of a worker with (1.17). In contrast, we consider here that the answers can be imprecise and affected by the experience of the workers. Hence, a novelty of our work is that the context of subjective answers is explicitly taken into account when determining the reliability of a worker.

7.3 Preliminaries

The aim of this section is to recall the key ideas for *modeling* and *comparing* the context of XBEs, as well as to present the problem statement and the notation used throughout the chapter.

7.3.1 Modeling Experience-Based Evaluations

As was mentioned in the introduction to this chapter, we consider that a subjective evaluation of a social media post could be given in a way that is general and not specific and, moreover, marked by hesitation. Aiming to model such XBEs, in the previous chapter we proposed an *augmented framework* that includes, among others, the definitions of *augmented appraisal degrees* and *augmented (Atanassov) intuitionistic fuzzy sets*. In what follows, we present some examples that show how these definitions can be applied.

7.3.1.1 Augmented Appraisal Degrees

When making an experience-based judgment of (the truth value of) a proposition, an evaluator could express his/her appraisal level along with something that hints the reasons of that appraisal. As has been explained in Chapter 6, an *augmented appraisal degree* [9] or *AAD* for short, is a generalization of a membership (or non-membership) grade [16, 17] that provides a (ready-for-computation) mathematical representation of this kind of judgments.

The necessity of such a mathematical representation can be illustrated as follows. Consider the proposition p : '*x is a social media post that supports the XYZ Act*', where x represents the social media post depicted in Figure 7.1. This proposition can be mathematically expressed in a canonical form 'x *IS* A' meaning 'x is an instance of A', where A is a (fuzzy) set of *'social media posts* *that support the XYZ Act'* [18]. Suppose that a requester and an anonymous worker focus on the *'drawing'* and the *'text'* features of x respectively, and their appraisals of p are *'quite not true'* and *'rather true'* in that order. If the answer of the worker is plainly expressed, i.e., it is like '*p is rather true*', it might mislead the requester into thinking that the worker is not doing a proper assessment. In contrast, if the worker's answer is '*p is rather true because of the "text" feature of x*', the requester might realize that the answer is valid from the worker's perspective. As could be noticed, a proper representation of such subjective evaluations is needed to avoid possible misunderstandings while processing them.

7.3.1.2 Augmented (Atanassov) Intuitionistic Fuzzy Sets

In evaluations resulting from a crowdsourcing request, an evaluator, say P, could judge the fulfillment of the criteria *"membership in A"* and *"non-membership in A"* on an object at the same time. Moreover, P could judge such criteria not for one, but for several objects within a collection, say X, according to his/her understanding of A. To manage this kind of subjective evaluation sets, in the previous chapter we proposed the inclusion of AADs into the definition of an *intuitionistic fuzzy set* (IFS) [16, 19]. Thereby, we illustrated that such an augmented IFS, called *augmented (Atanassov) intuitionistic fuzzy set* or AAIFS for short, can be used to model a collection of XBEs having judgments marked by hesitation. For example, consider the criteria "*membership in A*" and "*non-membership in A*" where A is (a collection of) *healthy sports*, and (the collection of sports) $X = \{$*tennis*, *football*$\}$. Consider also a unit interval scale where 1 denotes the highest level and 0 the lowest. Consider finally that a person, say P, makes the following judgments:

- *Tennis* fulfills with a grade of 0.7 the *"membership in healthy sports"* criterion because it is a non-contact sport and allows you to burn off some calories; however, due to the possibility of getting a tennis elbow, it fulfills with a grade of 0.1 the *"non-membership in healthy sports"* criterion as well.
- *Football* fulfills with a grade of 0.6 the *"membership in healthy sports"* criterion because it allows you to burn off some calories; in addition, it fulfills with a grade of 0.3 the *"non-membership in healthy sports"* criterion since the physical contact during a match could hurt you.

In this case, the judgments can be represented by an AAIFS, say $\hat{A}_{@P}$, such that

$$\begin{aligned}\hat{A}_{@P} = \{ \ &\big\langle \mathit{tennis}, \langle 0.7, \{\text{`}\mathit{non\text{-}contact}\text{'}, \text{`}\mathit{calories\ burn}\text{'}\}\rangle, \langle 0.1, \{\text{`}\mathit{tennis\ elbow}\text{'}\}\rangle \big\rangle, \\ &\big\langle \mathit{football}, \langle 0.6, \{\text{`}\mathit{calories\ burn}\text{'}\}\rangle, \langle 0.3, \{\text{`}\mathit{contact}\text{'}\}\rangle \big\rangle \ \} .\end{aligned}$$

Here, the hesitation of P to judge *tennis* as a member or not of *healthy sports* could be, e.g., $\hat{h}_{A@P}(\mathit{tennis}) = \langle 0.2, \{\text{`}\mathit{duration\ of\ a\ match}\text{'}\}\rangle$ – notice that the hesitation level results from $(1 - (0.7 - 0.1) = 0.2)$.

7.3.2 Comparing the Contexts of XBEs

As was mentioned in the previous part, characterizing XBEs as AADs can help to avoid misinterpretations when someone compares two of them. To illustrate this, let us consider the following request: using a unit interval scale where 1 represents the highest level and 0 the lowest, evaluate to which degree the comic book *'Popeye the Sailor'* is suitable for 7-year-old children. Two evaluators, say P and Q, both judge this comic with 0.4: while P does so because it contains some slang words and some illustrations of smokers and violence, Q assigns that level because he/she does not like spinach and, also, because of the slang words on it.

In this case, the judgments of the comic book $x =$ *'Popeye the Sailor'* can be characterized as the AADs $\hat{\mu}_{A@P}(x) = \langle 0.4, F_{\mu_A@P}(x)\rangle$ and $\hat{\mu}_{A@Q}(x) = \langle 0.4, F_{\mu_A@Q}(x)\rangle$ respectively, where A represents a collection of *comic books suitable for 7-year-old children*, and $F_{\mu_A@P}(x)$ and $F_{\mu_A@Q}(x)$ represent two collections of hints such that

$$F_{\mu_A@P}(x) = \{\textit{'slang expressions'}, \textit{'depiction of smokers'}, \\ \textit{'depiction of violence'}\}$$

and

$$F_{\mu_A@Q}(x) = \{\textit{'eating spinach'}, \textit{'slang expressions'}\}.$$

By comparing the collections of hints contained into these AADs, i.e., $F_{\mu_A@P}(x)$ and $F_{\mu_A@Q}(x)$, one can detect that P and Q have not focused on the same features of the comic book during their assessments. Hence, although the appraisal levels of these assessments match, one can realize that the assessments have some contextual differences. By the contrary, if only the appraisal levels are taken into account, i.e., $\mu_{A@P}(x) = 0.4$ and $\mu_{A@Q}(x) = 0.4$, one might assume that P and Q have focused on the same features of the comic book and, thus, wrongly state that both assessments match. As noticed, an advantage of this characterization is that it helps to compare the contexts of XBEs and, thus, to avoid this kind of '*pseudo-matching*'.

To quantify the comparison of the contexts in the previous case, in Section 6.4.1 we proposed a number $\Delta_{\mu_A} \in [0,1]$, named *connotation alikeness factor* (CAF), which indicates the level to which $F_{\mu_A@P}(x)$ and $F_{\mu_A@Q}(x)$ are perceived as similar.

Using a direct approach, P and Q can establish, e.g., $\Delta_{\mu_A@P} = 0.33$ and $\Delta_{\mu_A@Q} = 0.5$ respectively as indicators of their perceived similarity levels between $F_{\mu_A@P}(x)$ and $F_{\mu_A@Q}(x)$: while P detects one out of three features in common, Q detects one out of two. Notice that the similarity level perceived by P might not be equal to the level perceived by Q, i.e., the value of Δ_{μ_A} will depend on the perspective of either P or Q – this human behavior that can be present during the assessment of a similarity statement was studied by Tversky in [20].

Although the definition of a CAF is based on the collection of features contained in *two AADs* in relation to an object x, it can also be an indicator of the similarity of the features contained in *two collections of AADs* in relation to the elements in a collection X. For instance, consider a collection of comic books $X = \{x_1, \cdots, x_n\}$. Consider also two collections $\mathbf{F}_{\mu_A@P}$ and $\mathbf{F}_{\mu_A@Q}$ such that $\mathbf{F}_{\mu_A@P}(X) = F_{\mu_A@P}(x_1) \cup \cdots \cup F_{\mu_A@P}(x_n)$ and $\mathbf{F}_{\mu_A@Q}(X) = F_{\mu_A@Q}(x_1) \cup \cdots \cup F_{\mu_A@Q}(x_n)$, where $F_{\mu_A@P}(x_i)$ and $F_{\mu_A@Q}(x_i)$ are the collection of features focused on x_i by P and Q respectively. In this context, one can say that $\Delta_{\mu_A:P,Q@P}$ indicates how similar $\mathbf{F}_{\mu_A@P}$ and $\mathbf{F}_{\mu_A@Q}$ are according to the standpoint of P. Analogously, one can establish a CAF for each criterion in XBEs characterized as two AAIFSs: while $\Delta_{\mu_A:P,Q@P}$ will be an indicator of the perceived similarity level between $\mathbf{F}_{\mu_A@P}$ and $\mathbf{F}_{\mu_A@Q}$ according to the perspective of P, $\Delta_{\nu_A:P,Q@P}$ will indicate the perceived similarity level between $\mathbf{F}_{\nu_A@P}$ and $\mathbf{F}_{\nu_A@Q}$ according to the same perspective.

To compute an approximation of a CAF, one can use the *ratio model* presented in Section 6.2.1 to define a *σ-ratio*, say $\sigma_{\mathcal{C}:P,Q@P}$, such that

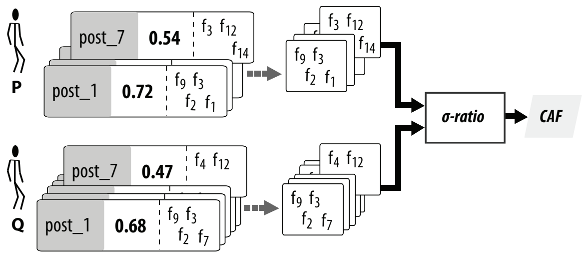
$$\sigma_{\mathcal{C}:P,Q@P} = \frac{n(F_{\mathcal{C}@P} \cap F_{\mathcal{C}@Q})}{n(F_{\mathcal{C}@P} \cap F_{\mathcal{C}@Q}) + \lambda_1 n(F_{\mathcal{C}@P} - F_{\mathcal{C}@Q}) + \lambda_2 n(F_{\mathcal{C}@Q} - F_{\mathcal{C}@P})} \tag{7.1}$$

is a measure of the similarity between the collections of hints $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$ as seen from the perspective of P. Analogous to (6.3), in this equation the expressions $n(F_{\mathcal{C}@P} \cap F_{\mathcal{C}@Q})$, $n(F_{\mathcal{C}@P} - F_{\mathcal{C}@Q})$ and $n(F_{\mathcal{C}@Q} - F_{\mathcal{C}@P})$ represent, in that order, the number of common hints, the number of hints that belong exclusively to $F_{\mathcal{C}@P}$ and the number of hints that belong exclusively to $F_{\mathcal{C}@Q}$. Likewise, $\lambda_1, \lambda_2 \in [0, 1]$ are parameters that adjust the contribution of $n(F_{\mathcal{C}@P} - F_{\mathcal{C}@Q})$ and $n(F_{\mathcal{C}@Q} - F_{\mathcal{C}@P})$ respectively – e.g., fixing values of $\lambda_1 = 1$ and $\lambda_2 = 0$ means that the hints belonging exclusively to $F_{\mathcal{C}@Q}$ are not taken into account to compute the ratio from P's perspective. As shown in Figure 7.2a, only the collections of hints given by two persons, say P and Q, are used by the σ-ratio to compute an approximation of a CAF.

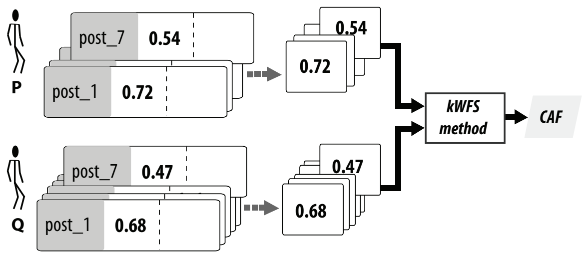
An alternative approach to compute such an approximation is related to the specific challenge presented in the introduction to this chapter, that is, *to make a computation using only the appraisal levels recorded in the AADs* (see Figure 7.2b). The problem dealt with in this approach is formulated as follows:

Problem statement: Consider a collection $X = \{x_1, \cdots, x_n\}$ and the criteria *"membership in A"* and *"non-membership in A"*, where A is a collection related to a particular concept. Consider also two persons, say P and Q. Let $\hat{A}_{@P}$ and $\hat{A}_{@Q}$ be two AAIFSs that represent the experience-based evaluations of the elements of X fulfilling the aforementioned criteria according to the perspectives of P and Q respectively; let $\Delta_{\mu_A:P,Q@P}$ and $\Delta_{\nu_A:P,Q@P}$ be the CAFs for the criteria *"membership in A"* and *"non-membership in A"* as seen from the perspective of P. Using the appraisal levels recorded in $\hat{A}_{@P}$ and $\hat{A}_{@Q}$, find two numbers in $[0, 1]$, say $\widetilde{\Delta}_{\mu_A:P,Q@P}$ and $\widetilde{\Delta}_{\nu_A:P,Q@P}$, such that they are the nearest approximations of $\Delta_{\mu_A:P,Q@P}$ and $\Delta_{\nu_A:P,Q@P}$ respectively.

In the next section, we describe the kWFS method, which aims to address this problem.



(a)



(b)

Figure 7.2: Two approaches to quantify the level to which the contexts of two collections of experience-based evaluations are alike.

7.4 k-Well-(Un)Fitted-Specimens Method

As was pointed out in the previous section, we are interested in computing an approximation of the perceived similarity level between the contexts of two collections of XBEs in which only the appraisal levels have been recorded. Hence, the question raised in this part is the following: *how to compute an approximation of a CAF by using not the hints but only the appraisal levels recorded in two collections of XBEs?*

To answer that question, we consider two ideas that originated after studying how individual experiences with a concept may conduct to different understandings of it (see Section 2.2), and how such different understandings are then reflected in a difference in the contexts of evaluations related to that concept (see Section 2.3). The first idea suggests that, during an evaluation process, one could detect some objects having features that make them well fitted (or unfitted) specimens of a concept. If a difference in understandings of the concept exists, it will be reflected through a difference between the appraisal levels of those objects. The second idea expresses that the appraisal levels recorded in evaluations given by persons whose understandings of a concept are in alignment, are expected to be more similar than the levels given by persons whose understandings are out of alignment. In what follows, we describe how these ideas, which were illustrated in Section 2.4, are reflected in the proposed kWFS method.

Algorithm 4: $\langle x_i, \hat{\mu}_{A@P}(x_i), \hat{\nu}_{A@P}(x_i)\rangle$ vs $\langle x_j, \hat{\mu}_{A@P}(x_j), \hat{\nu}_{A@P}(x_j)\rangle$

$h_{A@P}(x_i) \leftarrow 1 - \mu_{A@P}(x_i) - \nu_{A@P}(x_i)$
$h_{A@P}(x_j) \leftarrow 1 - \mu_{A@P}(x_j) - \nu_{A@P}(x_j)$
switch $focusedCriterion$ **do**
    case *'membership'* **do** `/* Block 1 */`
        if $\mu_{A@P}(x_i) > \mu_{A@P}(x_j)$ **then** $ret \leftarrow 1$
        else if $\mu_{A@P}(x_i) = \mu_{A@P}(x_j)$ **then**
            if $h_{A@P}(x_i) > h_{A@P}(x_j)$ **then** $ret \leftarrow 1$
            else if $h_{A@P}(x_i) = h_{A@P}(x_j)$ **then** $ret \leftarrow 0$
            else $ret \leftarrow -1$
        else $ret \leftarrow -1$
    case *'non-membership'* **do** `/* Block 2 */`
        if $\nu_{A@P}(x_i) > \nu_{A@P}(x_j)$ **then** $ret \leftarrow 1$
        else if $\nu_{A@P}(x_i) = \nu_{A@P}(x_j)$ **then**
            if $h_{A@P}(x_i) < h_{A@P}(x_j)$ **then** $ret \leftarrow 1$
            else if $h_{A@P}(x_i) = h_{A@P}(x_j)$ **then** $ret \leftarrow 0$
            else $ret \leftarrow -1$
        else $ret \leftarrow -1$
return ret

7.4.1 Identifying the k-Well-(un)Fitted Specimens

A possible way to automatically determine which of the evaluated objects can be considered well fitted (or unfitted) is to arrange them in descending order according to their appraisal levels and, then, choose the k objects at the top. To do so, one can use a sorting algorithm such as *QuickSort* [21] or *HeapSort* [22] with the comparison procedure described in Algorithm 4. This algorithm compares the appraisal levels of two objects, say x_i and x_j, given from the same perspective, say P, and returns 1 when x_i fits better than x_j, 0 when both x_i and x_j fit at the same level, and -1 when x_i fits worse than x_j.

As could be observed, Algorithm 4 performs a comparison according to which objects are needed: when well fitted specimens are needed, the "membership-in-A" levels are taken into account by executing the code in Block 1, and when unfitted specimens are needed, the algorithm uses the code in Block 2 to perform the comparison on the basis of the "non-membership-in-A" levels.

7.4.2 Identifying and Quantifying Proper Alignments

To identify by means of the appraisal levels given by two persons, say P and Q, whether their individual understandings about a concept, say A, are in or out of alignment, the second idea can be roughly translated to the following procedure:

(i) Consider that another person, say O, has an *"opposite-to-P"* understanding. Then, represent the appraisal levels of an object, say x_i, given by P, Q and O as *unit segments* constituted by $\mu_A(x_i)$, $\nu_A(x_i)$ and $h_A(x_i)$ such that $(\mu_A(x_i) + \alpha h_A(x_i)) + (\nu_A(x_i) + \beta h_A(x_i)) \leq 1$, $\alpha + \beta \leq 1$ and $\alpha, \beta \in [0, 1]$ (see Figures 7.3a and 7.3b) – here, α and β are deemed to be the *membership* and *non-membership hesitation splitters* respectively,

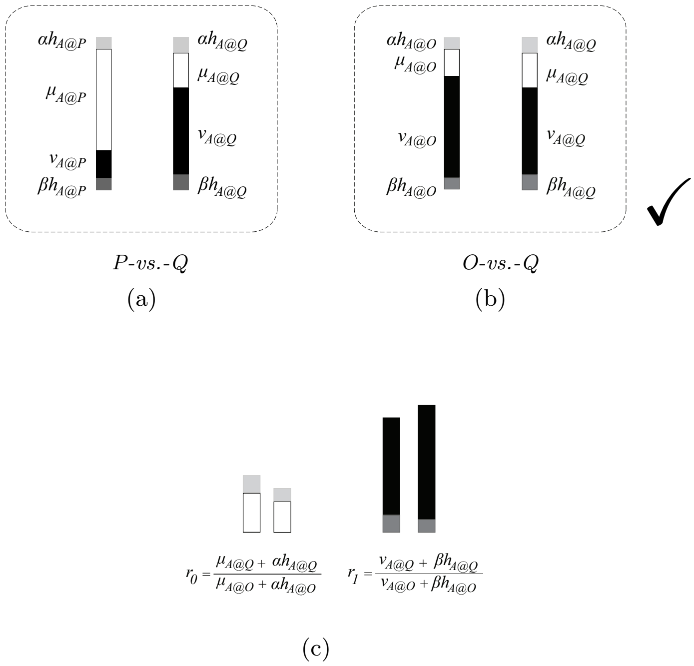


Figure 7.3: Identifying and quantifying proper alignments.

which split the hesitation level into the membership and non-membership parts according to a particular strategy (see Section 3.3.1).

(ii) Compare the unit segment corresponding to (the appraisal levels given by) Q to both the unit segment corresponding to P (see Figure 7.3a) and the unit segment corresponding to O (see Figure 7.3b). Assume that when the unit segments in *P-vs.-Q* are '*more similar*' than the ones in *O-vs.-Q*, P's and Q's understandings are rather deemed to be *in alignment*; and when the converse happens, P's and Q's understandings are rather deemed to be *out of alignment*. For instance, since the unit segments in Figure 7.3b (i.e., *O-vs.-Q*) are '*more similar*' than the unit segments in Figure 7.3a (i.e., *P-vs.-Q*), the individual understandings that O and Q have, are more in alignment than those of P and Q.

(iii) Assess to which degree P's and Q's understandings are in (or out of) alignment. This can be done by computing two ratios, say r_0 and r_1, of the appraisal levels of Q to the appraisal levels of either P or O depending on which of them is more in alignment with Q as indicated in the following equations:

$$r_0 = \frac{\mu_{A@Q} + \alpha h_{A@Q}}{\mu_{A@\{P|O\}} + \alpha h_{A@\{P|O\}}} \tag{7.2}$$

and

$$r_1 = \frac{\nu_{A@Q} + \beta h_{A@Q}}{\nu_{A@\{P|O\}} + \beta h_{A@\{P|O\}}}. \tag{7.3}$$

For instance, since we identified that O's understanding is more in alignment with Q's (compare Figure 7.3a and Figure 7.3b), we compute the ratios of the appraisal levels of Q to O with the equations

$$r_0 = \frac{\mu_{A@Q} + \alpha h_{A@Q}}{\mu_{A@O} + \alpha h_{A@O}}$$

and

$$r_1 = \frac{\nu_{A@Q} + \beta h_{A@Q}}{\nu_{A@O} + \beta h_{A@O}}$$

as shown in Figure 7.3c. It is worth mentioning that the aim of the ratios r_0 and r_1 is to detect how aligned with respect to the understanding of P, the understanding of Q is. Hence, we assume that the appraisal levels of the k relevant objects identified by P are always available, i.e., we assume that the AADs of the k relevant objects identified by P are at least plain AADs. However, *missing appraisals* of such relevant objects, i.e. appraisals in which neither the levels nor the hints are available, are possible in the case of Q. If that is the case, the *default* appraisal levels, i.e., $\mu_{A@Q} = 0$, $\nu_{A@Q} = 0$ and $h_{A@Q} = 1$, will be used for the computation of r_0 and r_1.

The above informally described steps are concretized in our proposed *k-well-(un)fitted-specimens (kWFS) method.* Algorithm 5 describes this method. The algorithm obtains an approximation of a CAF for $\hat{A}_{@P}$-vs-$\hat{A}_{@Q}$ from P's perspective, where $\hat{A}_{@P}$ and $\hat{A}_{@Q}$ are two AAIFSs representing the experience-based evaluation sets given by P and Q respectively.

To begin with, the objects in X are arranged in descending order depending on which CAF ($\Delta_{\mu_A:P,Q@P}$ or $\Delta_{\nu_A:P,Q@P}$) is being approximated to (Lines 1-4). To do so, the method $sortByMembership$ (Line 2) makes use of Block 1 in Algorithm 4; while the method $sortByNonMembership$ (Line 4) makes use of Block 2.

Then, a *spot CAF* (identified by caf_i) is computed for each of the top k objects in the ordered collection $X^{'}$ as follows:

(i) In Line 8, caf_i is initialized to 0.5 for the sake of neutrality (recall from Definition 6.2 that 1 denotes the highest level of similarity between the focused features and 0 represents the lowest).
(ii) In Line 9, a spot difference (identified by dif_{given}) between the given appraisal levels is computed by means of (3.9), i.e.,

$$\begin{aligned}\mathrm{dif}^{\alpha,\beta}(\mathbf{p_i}, \mathbf{q_i}) = (\mu_{A@P}(x_i) + \alpha h_{A@P}(x_i))(\nu_{A@Q}(x_i) + \beta h_{A@Q}(x_i)) \\ - (\nu_{A@P}(x_i) + \beta h_{A@P}(x_i))(\mu_{A@Q}(x_i) + \alpha h_{A@Q}(x_i)),\end{aligned}$$

where

$$\mathbf{p_i} = \begin{pmatrix} \mu_{A@P}(x_i) & + & \alpha h_{A@P}(x_i) \\ \nu_{A@P}(x_i) & + & \beta h_{A@P}(x_i) \end{pmatrix}$$

Algorithm 5: k-well-(un)fitted-specimens Method.

Data: X, $\hat{A}_{@P}$, $\hat{A}_{@Q}$, k, α, β, $cafType$
Result: Approximation of a CAF for $\hat{A}_{@P}$-vs-$\hat{A}_{@Q}$ from P's perspective

1 **if** $cafType$ = *'membership'* **then**
2 $\quad X' \leftarrow X$.`sortByMembership(`$\hat{A}_{@P}$,*descending*`)`
3 **else**
4 $\quad X' \leftarrow X$.`sortByNonMembership(`$\hat{A}_{@P}$,*descending*`)`
5 $c \leftarrow 0$
6 $caf \leftarrow 0$
7 **foreach** $x_i \in X'$ **do**
8 $\quad r_0 \leftarrow 0.5, r_1 \leftarrow 0.5, caf_i \leftarrow 0.5$ `/* neutrality */`
9 $\quad dif_{given} \leftarrow (\mu_{A@P}(x_i) + \alpha h_{A@P}(x_i))(\nu_{A@Q}(x_i) + \beta h_{A@Q}(x_i)) - (\nu_{A@P}(x_i) + \beta h_{A@P}(x_i))(\mu_{A@Q}(x_i) + \alpha h_{A@Q}(x_i))$
10 $\quad dif_{opposite} \leftarrow (\nu_{A@P}(x_i) + \beta h_{A@P}(x_i))(\nu_{A@Q}(x_i) + \beta h_{A@Q}(x_i)) - (\mu_{A@P}(x_i) + \alpha h_{A@P}(x_i))(\mu_{A@Q}(x_i) + \alpha h_{A@Q}(x_i))$
11 $\quad$ **if** $|dif_{opposite}| < |dif_{given}|$ **then**
12 $\quad\quad$ **if** $\nu_{A@P}(x_i) + \beta h_{A@P}(x_i) > 0$ **then**
13 $\quad\quad\quad r_0 \leftarrow (\mu_{A@Q}(x_i) + \alpha h_{A@Q}(x_i))/(\nu_{A@P}(x_i) + \beta h_{A@P}(x_i))$
14 $\quad\quad$ **if** $\mu_{A@P}(x_i) + \alpha h_{A@P}(x_i) > 0$ **then**
15 $\quad\quad\quad r_1 \leftarrow (\nu_{A@Q}(x_i) + \beta h_{A@Q}(x_i))/(\mu_{A@P}(x_i) + \alpha h_{A@P}(x_i))$
16 $\quad\quad caf_i \leftarrow 1 - \min(1, \max(r_0, r_1))$
17 $\quad$ **else if** $|dif_{opposite}| > |dif_{given}|$ **then**
18 $\quad\quad$ **if** $\mu_{A@P}(x_i) + \alpha h_{A@P}(x_i) > 0$ **then**
19 $\quad\quad\quad r_0 \leftarrow (\mu_{A@Q}(x_i) + \alpha h_{A@Q}(x_i))/(\mu_{A@P}(x_i) + \alpha h_{A@P}(x_i))$
20 $\quad\quad$ **if** $\nu_{A@P}(x_i) + \beta h_{A@P}(x_i) > 0$ **then**
21 $\quad\quad\quad r_1 \leftarrow (\nu_{A@Q}(x_i) + \beta h_{A@Q}(x_i))/(\nu_{A@P}(x_i) + \beta h_{A@P}(x_i))$
22 $\quad\quad caf_i \leftarrow \min(1, \max(r_0, r_1))$
23 $\quad caf \leftarrow caf + caf_i$
24 $\quad c \leftarrow c + 1$
25 $\quad$ **if** $c = k$ **then return** caf/k

and

$$\mathbf{q_i} = \begin{pmatrix} \mu_{A@Q}(x_i) & + & \alpha h_{A@Q}(x_i) \\ \nu_{A@Q}(x_i) & + & \beta h_{A@Q}(x_i) \end{pmatrix}$$

are the vector interpretations of the appraisal levels given by P and Q respectively according to (3.7).

(iii) In Line 10, an "opposite" spot difference (identified by $dif_{opposite}$) between the given appraisal levels is computed by means of the equation

$$\begin{aligned}\mathrm{dif}^{\alpha,\beta}(\mathbf{o_i}, \mathbf{q_i}) = {} & (\nu_{A@P}(x_i) + \beta h_{A@P}(x_i))(\nu_{A@Q}(x_i) + \beta h_{A@Q}(x_i)) \\ & - (\mu_{A@P}(x_i) + \alpha h_{A@P}(x_i))(\mu_{A@Q}(x_i) + \alpha h_{A@Q}(x_i)),\end{aligned}$$

where

$$\mathbf{o_i} = \begin{pmatrix} \nu_{A@P}(x_i) & + & \beta h_{A@P}(x_i) \\ \mu_{A@P}(x_i) & + & \alpha h_{A@P}(x_i) \end{pmatrix}$$

is the vector interpretation of the appraisal levels as given according to an understanding that is opposite to P's.

(iv) In Line 11, the *"out-of-alignment"* condition is tested. If this condition is satisfied, the *membership ratio* (identified by r_0) between the membership levels in $\mathbf{q_i}$ and $\mathbf{o_i}$ is computed in Line 13, as well as the *non-membership ratio* (identified by r_1) between their corresponding non-membership levels is computed in Line 15. In this case, the value of caf_i results from the difference between 1 (i.e., the highest possible value) and the highest ratio of r_0 and r_1 (Line 16).

(v) In Line 17, the *"in-alignment"* condition is tested. When this condition is satisfied, the *membership ratio* (identified by r_0) of the membership levels in $\mathbf{q_i}$ to the the membership levels in $\mathbf{p_i}$ is computed in Line 19, as well as the *non-membership ratio* (identified by r_1) between their corresponding non-membership levels is computed in Line 21. In this case, the value of caf_i is set to the highest ratio of r_0 and r_1 (Line 22).

(vi) If neither of the above conditions holds (i.e., if $dif_{given} = dif_{opposite}$), then the value of caf_i is kept as initialized in Line 8 (i.e., $caf_i = 0.5$).

After computing the k spot CAFs, their overall average is calculated and returned as a resulting approximation of the required CAF (Line 25). In the next section, an empirical procedure that was performed to test the effectiveness of our method is described.

7.5 Testing the kWFS Method

As was mentioned in Section 7.3, when the reasons of a judgment are recorded in two collections of XBEs, one can quantify the perceived similarity between their contexts using an approach in which a CAF is directly chosen from a range of values. Also, it is possible to use a psychological approach in which an approximation of a CAF is computed by the σ-ratio (see (7.1)).

The proposed kWFS method computes such an approximation by using only the appraisal levels recorded in the XBEs. Hence, to obtain a good evidence of its effectiveness, one can compare the approximations resulting from the kWFS method with the approximations resulting after applying the approaches mentioned in Section 7.3.2. Furthermore, these approximations can be employed to detect evaluators who can perform a job according to a particular understanding. Thus, the results can be also compared with the reliability levels computed with methods that use a gold standard collection (see Section 7.2).

To evaluate the kWFS method, we opt to use the XBEs resulting from the simulated experience-based evaluation process described in Section 4.3. A justifiable reason for using such XBEs is that one can obtain good insights about how the perceive similarity between the contexts is affected by what the evaluators have learned, as well as how the kWFS method can identify people with different (or similar) understandings.

In the first part of this section, we briefly recall how those simulated XBEs were obtained. Then, in the second part, we use the XBEs to determine the effectiveness of the kWFS method while detecting people having similar (or different) understandings. Finally, the results and a discussion about them are presented.

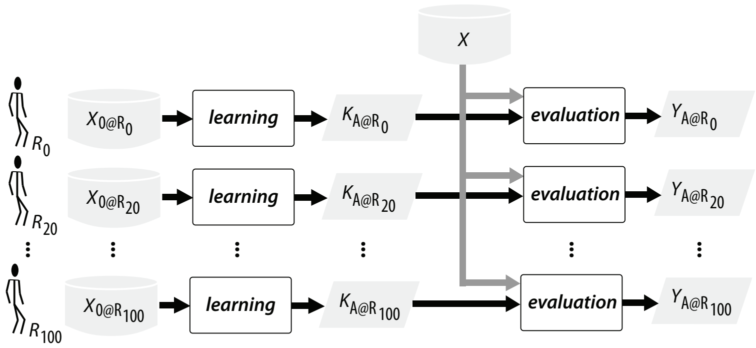


Figure 7.4: General view of a process by which experience-based evaluations are obtained.

7.5.1 Obtaining Experience-Based Evaluations

To obtain XBEs, a simulated process in which evaluations conducted by people who learned how to categorize newswire stories is used (see Figure 7.4). In this process, a person, say R_{20}, is associated to a scenario in which a *particular training data set*, say $X_{0@R_{20}}$, is used as an input of a *learning process* to learn about newswire stories belonging to a given category, say A – as was done in Section 2.2.3, we use the subscript '0@P' to denote a training data set used by a person P. At the end of the learning process, a *knowledge model*, say $K_{A@R_{20}}$, is obtained. Then, the knowledge model $K_{A@R_{20}}$ along with a *test data set*, say X, are used as inputs of an *evaluation process* to obtain a collection of XBEs, say $Y_{A@R_{20}}$. After conducting the learning process, a knowledge representation $K_{A@P}$ was obtained for each category A in {*ECAT, E11, E12, GSCI, GSPO, GTOUR, GVIO, CCAT, C12, C13, GCAT, G15, GDEF, GDIP, GDIS, GENT, GENV, GFAS, GHEA, GJOB*} learned by each person P in $\{R_0, R_{20}, R_{40}, R_{60}, R_{80}, R_{100}\}$.

To simulate a qualification process in which an anonymous potential collaborator is asked to evaluate a collection of newswire stories in order to demonstrate his/her understanding about several categories, the collections of XBEs were obtained as follows. Consider a collection X having 50 stories. First, to evaluate the level to which each story in X fits into a category, say *G15*, according to the individual understanding that a person, say R_{80}, has about this category, the corresponding knowledge representation, i.e., $K_{G15@R_{80}}$, is used in the evaluation process described in Section 2.3. After evaluating all the stories in X, the collection $Y_{G15@R80}$, which represents the resulting collection of XBEs in this case, is built. These steps are repeated for each X in $\mathcal{X}$, where $\mathcal{X}$ represents the 1000 50-story collections built for the test.

Recall that in Section 4.3 we chose 20 categories and 1000 50-story collections and, thus, 20000 XBE sets were obtained for each person. This means that a total of 120000 XBEs can be used as a (secondary) data set to determine the effectiveness of the kWFS method while detecting people with similar understandings. So is done in the next part.

7.5.2 Detecting People Having Similar Understandings

As was mentioned in the introduction to this chapter, we propose to detect people having similar understandings by comparing the contexts of their XBEs. Since a similarity comparison between the contexts of two collections of XBEs can be made through a CAF (see Section 7.3.2), we employ the kWFS method and the σ-ratio to compute approximations of CAFs between the collection of XBEs given by a referent evaluator, say R_0, and the collection of XBEs given by the other evaluators. To do so, we represent the collection of XBEs by AAIFSs as follows:

Let $\mathcal{C}$ be a criterion having a form such as *"membership in A,"* where A is a given concept; let X be a test data set that is part of an evaluation request; and let x_i be one of the objects in X. Consider a particular part of knowledge about A, say K_A, which is characterized by $\langle \hat{\mathbf{u}}_A, t_A \rangle$, where $\hat{\mathbf{u}}_A = \omega_1 \hat{\mathbf{f}}_1 + \cdots + \omega_m \hat{\mathbf{f}}_m$, and $\hat{\mathbf{f}}_j$ is a unit vector that represents the dimension corresponding to the feature f_j (see Section 2.2.2). Consider also a vector $\mathbf{f}_{i,j} = \beta_{i,j} \hat{\mathbf{f}}_j$, which represents the *overall influence* of f_j when the fulfillment of $\mathcal{C}$ is appraised on x_i. Finally, consider a vector $\mathbf{x}_i = \beta_{i,1} \hat{\mathbf{f}}_1 + \cdots + \beta_{i,m} \hat{\mathbf{f}}_m$, which represents the *resulting overall influence* of the features of x_i on such an appraisal. With these assumptions, a procedure for representing an evaluation as an AAIFS element consists of the following two steps:

1. Compute the *specific influence* of a feature f_j (i.e., $\mathbf{f}_{i,j_A} = \beta_{i,j_A} \hat{\mathbf{u}}_A$, where $\beta_{i,j_A} = \beta_{i,j}\omega_j$) to identify whether f_j has been focused on during the judgment of "x_i satisfies $\mathcal{C}$": if $\beta_{i,j_A} = 0$, f_j has not been focused on; if $\beta_{i,j_A} > 0$, f_j has been focused on and favors the evaluation criterion, i.e., the direction of $\mathbf{f}_{i,j_A}$ is the same as the direction of $\hat{\mathbf{u}}_A$; if $\beta_{i,j_A} < 0$, f_j has been focused on and disfavors the evaluation criterion, i.e., the direction of $\mathbf{f}_{i,j_A}$ is opposite to the direction of $\hat{\mathbf{u}}_A$.
2. Compute $\mu_A(x_i)$ and $\nu_A(x_i)$ by means of the equations

$$\mu_A(x_i) = \check{\mu}_A(x_i)/\eta \tag{7.4}$$

and

$$\nu_A(x_i) = \check{\nu}_A(x_i)/\eta \tag{7.5}$$

respectively, where

$$\check{\mu}_A(x_i) = \begin{cases} \frac{1}{\|\mathbf{x}_i\|}\left(|t_A| + \sum_{j=1}^{m} \beta_{i,j_A}\right) & \text{if } \left(\forall j : \beta_{i,j_A} > 0\right) \wedge (t_A < 0)\,; \\ \frac{1}{\|\mathbf{x}_i\|}\left(\sum_{j=1}^{m} \beta_{i,j_A}\right) & \text{if } \left(\forall j : \beta_{i,j_A} > 0\right) \wedge (t_A \geq 0)\,; \\ 0 & \text{otherwise;} \end{cases} \tag{7.6}$$

$$\check{\nu}_A(x_i) = \begin{cases} \frac{1}{\|\mathbf{x}_i\|}\left(t_A + \sum_{j=1}^{m} |\beta_{i,j_A}|\right) & \text{if } \left(\forall j : \beta_{i,j_A} < 0\right) \wedge (t_A > 0) \\ \frac{1}{\|\mathbf{x}_i\|}\left(\sum_{j=1}^{m} |\beta_{i,j_A}|\right) & \text{if } \left(\forall j : \beta_{i,j_A} < 0\right) \wedge (t_A \leq 0)\,; \\ 0 & \text{otherwise;} \end{cases} \tag{7.7}$$

and

$$\eta = \max\left(1, \check{\mu}_A(x_i) + \check{\nu}_A(x_i)\right), \forall x_i \in X. \tag{7.8}$$

Notice that the above equations are analogous to (4.11), (4.12), (4.13), (4.14) and (4.15) respectively.

Using the resulting AAIFS, we compute the approximations of CAFs with the kWFS method (Algorithm 5) and the σ-ratio (see (7.1)). Then, we aggregate the obtained approximations through the averages

$$\overline{\Delta}_{\mathcal{C}:P,Q@P}(\mathcal{X}) = \frac{1}{n(\mathcal{X})} \sum_{X \in \mathcal{X}} \widetilde{\Delta}_{\mathcal{C}:P,Q@P}(X) \tag{7.9}$$

and

$$\overline{\sigma}_{\mathcal{C}:P,Q@P}(\mathcal{X}) = \frac{1}{n(\mathcal{X})} \sum_{X \in \mathcal{X}} \sigma_{\mathcal{C}:P,Q@P}(X) \tag{7.10}$$

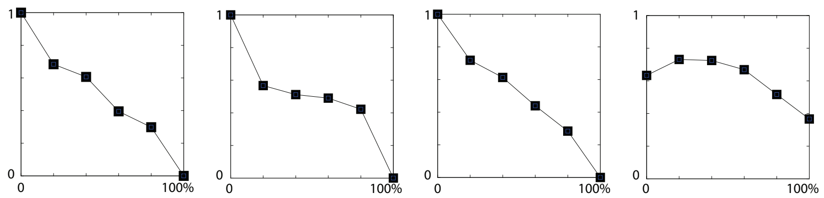
respectively, where $n(\mathcal{X})$ represents the number of collections in $\mathcal{X}$. The effectiveness of both methods will be determined based on how well those averages identify the people with similar understandings. We can expect that the average CAF obtained from the collection of XBEs given by people having similar experiences will be greater than the average CAF obtained from the collection of XBEs given by people with different experiences. For instance, since the training collections used by persons R_0 and R_{20} while learning about category $G15$ are more similar than the ones used by R_0 and R_{100}, we can expect that $\overline{\Delta}_{\mu_{G15}:R_0,R_{20}@R_0}(\mathcal{X}) > \overline{\Delta}_{\mu_{G15}:R_0,R_{100}@R_0}(\mathcal{X})$ will hold. We can also expect that $\overline{\Delta}_{\mu_{G15}:R_0,R_0@R_0}(\mathcal{X})$ and $\overline{\Delta}_{\mu_{G15}:R_0,R_{100}@R_0}(\mathcal{X})$ will have the highest and the lowest values respectively.

To identify (reliable) workers having similar understandings by means of the methods that use a gold standard collection, we characterize the XBEs that result from these methods as '*crisp*' evaluations, i.e., evaluations in which the stories have full membership (or nonmembership) in a category. To do so, we assume that a story is full member of a category if the level to which the story belongs to the category is greater than the level to which the story does not belong to that category, and vice versa. Then, we use two strategies to build gold standard collections of XBEs: one in which the XBEs of newswire stories randomly selected by the person who is referent (i.e., R_0) are chosen, and the other in which *collective XBEs*, i.e., XBEs that result from the aggregation of the XBEs given by all the evaluators, are deemed to be part of the gold standard collection. After that, we use (1.17) to compute the reliability of each worker using the gold standard collections resulting from the aforementioned strategies by means of the equations

$$s_{f.gold}(X) = \frac{g(X)}{G_f(X)} \tag{7.11}$$

and

$$s_{c.gold}(X) = \frac{g(X)}{G_c(X)}, \tag{7.12}$$



(a) $\overline{\Delta}_{\mu}$-0.5-0.5-5 (b) $\overline{\sigma}_{\mu}$-1-0 (c) $\overline{s}_{f.gold}$-5 (d) $\overline{s}_{c.gold}$-5

Figure 7.5: Averages of the computed CAFs and reliability scores versus the percentage of opposites included in each scenario.

where $g(X)$ denotes the number of 'correct' XBEs (in this case correct categorizations), and $G_f(X)$ and $G_c(X)$ represent, respectively, the size of the gold standard collection with randomly selected newswire stories and the size of the gold standard collection with collective XBEs. Finally, as was done with the approximations of CAFs computed by the kWFS method and the σ-ratio, the reliability scores resulting from the fixed and collective strategies are aggregated by computing arithmetic averages, i.e.,

$$\overline{s}_{f.gold} = \frac{1}{n(\mathcal{X})} \sum_{X \in \mathcal{X}} s_{f.gold}(X) \tag{7.13}$$

and

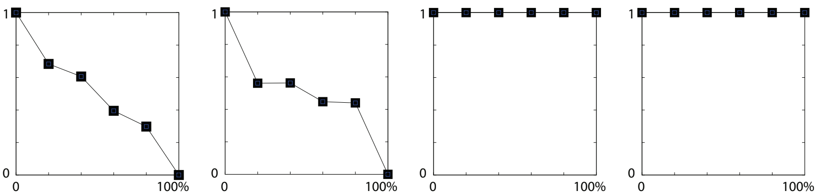
$$\overline{s}_{c.gold} = \frac{1}{n(\mathcal{X})} \sum_{X \in \mathcal{X}} s_{c.gold}(X) \tag{7.14}$$

respectively. The results of these aggregations are presented in the next part.

7.5.3 Results

The results of the averages of the CAFs computed by the kWFS method and the σ-ratio versus the percentage of opposites included in each scenario are depicted in Figures 7.5a and 7.5b respectively. While Figure 7.5a is captioned with the nomenclature $\overline{\Delta}_{\mu}$-α-β-k to denote the values of the parameters selected for the kWFS method, Figure 7.5b is captioned with $\overline{\sigma}_{\mu}$-λ_1-λ_2 to denote so for the σ-ratio. As can be noticed, the averages of the CAFs computed by these (configurations of the) methods decrease with the increment of opposites. This means that, in average (a proper configuration of) these methods can identify how similar the contexts of the simulated XBEs are in relation to the particular learning experiences of their providers. For instance, since $\overline{\Delta}_{\mu:R_0,R_{20}@R_0} > \overline{\Delta}_{\mu:R_0,R_{60}@R_0}$ holds according to Figure 7.5a, we can say that, compared to R_{20}'s and R_{60}'s understandings, the understanding of R_0 is more similar to R_{20}'s.

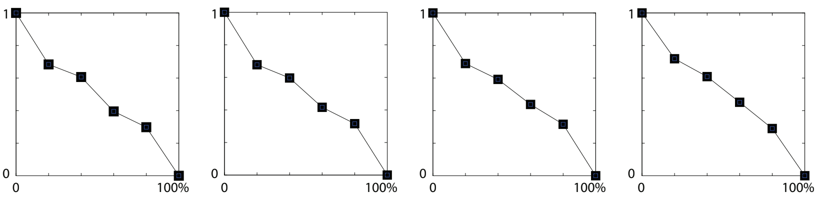
Regarding the reliability scores, the results of the averages $\overline{s}_{f.gold}$ and $\overline{s}_{c.gold}$ versus the percentage of opposites are shown in Figures 7.5c and 7.5d respectively. The figures are captioned $\overline{s}_{f.gold}$-k and $\overline{s}_{c.gold}$-k to indicate that gold



(a) $\overline{\Delta}_{\mu}$-0.5-0.5-k (b) $\overline{\Delta}_{\mu}$-0.5-0.5-k (c) $\overline{\Delta}_{\mu}$-0-1-k (d) $\overline{\Delta}_{\mu}$-1-0-k

Figure 7.6: Variation on parameters α and β of the kWFS method ($k = 5$).

collections with k elements have been used. As noticed, while the averages of the $\overline{s}_{f.gold}$ scores strictly decrease with the increment of opposites, the averages of the $\overline{s}_{c.gold}$ scores increase first and then decrease with the increment of opposites. This suggests that, on average, only the former score can identify through the simulated 'crisp' XBEs workers who are more reliable when performing a job on behalf of the referent R_0 – where, someone who shares a similar understanding with R_0 is deemed to be a reliable worker. For the same reason, the results also suggest that the average $\overline{s}_{c.gold}$ might identify the workers who are more reliable in relation to a collective understanding resulting from heterogeneous learning experiences.



(a) $\overline{\Delta}_{\mu}$-α-β-5 (b) $\overline{\Delta}_{\mu}$-α-β-10 (c) $\overline{\Delta}_{\mu}$-α-β-25 (d) $\overline{\Delta}_{\mu}$-α-β-50

Figure 7.7: Variation on parameter k of the kWFS method ($\alpha = 0.5$ and $\beta = 0.5$).

To observe how the results can be affected by the variation on the parameters of the kWFS method, several configurations were tested. These results are shown in Figures 7.6 and 7.7: while the former figure shows how the results are affected by a variation of the parameters α and β, the latter figure presents the results when the parameter k is changed. Notice that, only when the hesitation splitters α and β are equally distributed or zeroed, the averages of the CAFs decrease with the increment of opposites. Notice also that the variation on the parameter k produces slight variations on the averages of the CAFs.

7.5.4 Discussion

The results reported above suggest that the proposed kWFS method can obtain a reasonable approximation of the CAFs between the simulated XBEs when it is configured with hesitation splitters that are equally distributed or zeroed. This means that, on average, our method can compute approximations of CAFs that are in agreement with approximations computed by applying the *sigma*-ratio, i.e., approximations resulting after processing the collection of features recorded in the XBEs. Thus, it can be used to identify people having similar understandings in the simulated process. In this respect, the results of the kWFS method are comparable with the results obtained after computing reliability scores that use a gold standard collection containing a fixed number of 'crisp' evaluations given by a referent person.

A possible explanation for the need of using hesitation splitters that are equally distributed or zeroed is that, due to the relative high number of features in the evaluated objects (i.e., the words in the evaluated newswire stories), the hesitation margin is so considerable that the given appraisal levels (i.e., the membership and non-membership levels) might be disregarded if such a margin is not equally distributed nor zeroed. Bearing in mind that such "high hesitation" XBEs are possible in a crowd-sourced environment, their proper handling could be seen as an advantage of the kWFS method – the interested reader may recall the results of the empirical study in Section 4.5 that shows how some similarity measures for IFSs failed to handle that kind of evaluations.

With respect to the way of getting simulated XBEs, establishing learning scenarios that contain a certain proportion of opposite examples in relation to the original data allows us to observe how the kWFS method can deal with the comparison of XBEs resulting from dissimilar learning experiences. For instance, we can observe that both the σ-ratio and the kWFS method with $\alpha = 0$ and $\beta = 0$ handle properly the comparison of the evaluation sets given by (the evaluators) *R0* and *R100*, who use training collections having examples that are totally opposite to each other. In addition, using XBEs resulting from such simulated learning experiences allows us to build situations in which flat comparisons could be problematic if not detected. Since XBEs resulting from extremely dissimilar learning experiences were used during the test, we were able to observe that our method with hesitation splitters $\alpha = 1$ and $\beta = 0$ does not compute appropriate approximations of CAFs. Hence, we can strongly suggest avoiding such parameters values if accuracy is important. For instance, imagine what would happen if these values are used for clustering XBEs according to the evaluator's understanding of any of the tested categories and the XBEs given by (the evaluators) *R0* and *R100* were put into the same group.

It is worth mentioning that, since the reported results are based on simulated XBEs, they should be only treated as a good indication of the effectiveness of the proposed method. Hence, conducting experiments with real evaluators is recommended and subject to further study.

7.6 Potential Applications

As was mentioned in Chapter 6, when a CAF is taken into consideration, a comparison between any two XBEs can be a reliable act even if the XBEs are fairly subjective. This means that, since the proposed method can obtain a reasonable approximation of a CAF between any pair of XBEs, it can potentially be applied for improving tasks such as querying, filtering, sorting or clustering where comparisons of subjective data are required. In this way, for example, personnel recruitment, content rating of advertisements or decision processes about subjective business proposals will benefit from the techniques described in this chapter.

7.7 Conclusions

In this chapter we have described a method that computes an approximation of the level to which the contexts of *experience-based evaluations* (XBEs) are alike. Such an approximation is deemed to be an indicator of what a person perceives as quality on subjective (fuzzy) judgments (Research Question $Q3$). Hence, it can be applied to detect whether the context of crowdsourced evaluations on social media content are in line with the context of the evaluations given by a requester (Research Questions $Q4$ and $Q5$).

In contrast to a direct approach in which the aforementioned indicator is fixed after studying the hints recorded in the evaluations under analysis, a novelty of the proposed method is that it relies only on the appraisal levels corresponding to a specific number of representative evaluations. An advantage on this new approach is that the method is also applicable to situations in which none or only a small number of hints are recorded in all (or a big number of) the evaluations under analysis – i.e., the method can be applied for comparing *flat XBEs*, in which only appraisal levels are available.

The results obtained after conducting an empirical test with simulated XBEs provide an adequate evidence of the effectiveness of the proposed method while identifying and measuring a potential difference in the contexts of the evaluations given by evaluators with different backgrounds. It was found that our method with a proper parameter configuration can obtain a good approximation of the similarity between the contexts of "high hesitation" XBEs (i.e., evaluations in which the level of hesitation is much greater than the sum of the appraisal levels). Bearing in mind that such "high hesitation" XBEs are possible in crowdsourcing, their proper handling can be seen as another potential advantage of our method. However, since the results reported are based on simulated XBEs, conducting experiments with real evaluators is recommended and subject to further study.

In the next chapter, we will explain how to use the approximations of the levels to which the contexts of *unrequested* XBEs are alike to detect people (or information sources) that reflect a similar understanding about a given topic.

References

[1] Andrea L. Kavanaugh, Edward A. Fox, Steven D. Sheetz, Seungwon Yang, Lin Tzy Li, Donald J. Shoemaker, Apostol Natsev, and Lexing Xie. *Social media use by government: From the routine to the critical.* Government Information Quarterly, 29(4):480 – 491, 2012.

[2] Ines Mergel. *Social media adoption and resulting tactics in the U.S. federal government.* Government Information Quarterly, 30(2):123 – 130, 2013.

[3] Daren C. Brabham. *Crowdsourcing.* Mit Press, 2013.

[4] Enrique Estellés-Arolas and Fernando González-Ladrón-de Guevara. *Towards an integrated crowdsourcing definition.* Journal of Information science, 38(2):189–200, 2012.

[5] Jeff Howe. *The Rise of Crowdsourcing.* Wired. Available at http://archive.wired.com/wired/archive/14.06/crowds.html, 2006. Date last accessed: October, 30th 2016.

[6] David Oleson, Alexander Sorokin, Greg P. Laughlin, Vaughn Hester, John Le, and Lukas Biewald. *Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing.* Human computation, 11(11), 2011.

[7] Petros Venetis and Hector Garcia-Molina. *Quality control for comparison microtasks.* In Proceedings of the first international workshop on crowdsourcing and data mining, pages 15–21. ACM, 2012.

[8] Stephanie Watts, Ganesan Shankaranarayanan, and Adir Even. *Data quality assessment in context: A cognitive perspective.* Decision Support Systems, 48(1):202–211, 2009.

[9] Marcelo Loor and Guy De Tré. *On the need for augmented appraisal degrees to manage experience-based evaluations.* Applied Soft Computing, 54:284 – 295, 2017.

[10] Vikas C. Raykar and Shipeng Yu. *Eliminating spammers and ranking annotators for crowdsourced labeling tasks.* Journal of Machine Learning Research, 13(Feb):491–518, 2012.

[11] Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Y. Ng. *Cheap and Fast—but is It Good?: Evaluating Non-expert Annotations for Natural Language Tasks.* In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 254–263. Association for Computational Linguistics, 2008.

[12] Sujoy Chatterjee and Malay Bhattacharyya. *Judgment analysis of crowd-sourced opinions using biclustering.* Information Sciences, 375:138–154, 2017.

[13] Dirk Hovy, Taylor Berg-Kirkpatrick, Ashish Vaswani, and Eduard H Hovy. *Learning whom to trust with MACE.* In HLT-NAACL, pages 1120–1130, 2013.

[14] Panagiotis G. Ipeirotis, Foster Provost, and Jing Wang. *Quality management on Amazon Mechanical Turk.* In Proceedings of the ACM SIGKDD workshop on human computation, pages 64–67. ACM, 2010.

[15] Chris Callison-Burch. *Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk.* In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1, EMNLP '09, pages 286–295, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[16] Krassimir T. Atanassov. *Intuitionistic fuzzy sets.* Fuzzy sets and Systems, 20(1):87–96, 1986.

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

[18] Lotfi A. Zadeh. *Is there a need for fuzzy logic?* Information Sciences, 178(13):2751 – 2779, 2008.

[19] Krassimir T. Atanassov. *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[20] Amos Tversky. *Features of similarity.* Psychological review, 84(4):327, 1977.

[21] C. A. R. Hoare. *Quicksort.* The Computer Journal, 5(1):10–16, 1962.

[22] John William Joseph Williams. *Algorithm-232-heapsort.* Communications of the ACM, 7(6):347–348, 1964.

Chapter 8

Handling Unrequested Experience-Based Evaluations

Abstract

As was mentioned in Chapter 1, experience-based evaluations (XBEs) can be a consequence of unsolicited opinions or judgments. This is the case for opinions posted on social media. Since it can result in significant benefits for a company or an individual, the inclusion of information extracted from such opinions into a decision making process is becoming a more frequent activity. However, such benefits are usually linked to the reliability of the extracted information, which, among other aspects, depends on the reliability of its source. In this regard, the question *"how to determine the reliability of a person who, without being explicitly asked for them, publishes on social media his/her opinions about a given topic?"* is raised. Aiming to address this question, in this chapter we propose a novel technique for handling *unrequested* XBEs. With this technique, posts on social media are digested to build a kind of database consisting of *augmented (Atanassov) intuitionistic fuzzy sets*, or AAIFSs for short, each resembling a collection of XBEs given by a particular source with respect to a topic under analysis. As explained in Chapter 6, such AAIFSs can be used in comparisons in which not only the judgments, but also their contexts are taken into account for computation. Hence, extracting more reliable information is possible. An illustrative example shows how the proposed technique works and how it can help to detect sources having a common understanding of a given topic.

This chapter is an adapted version of the following manuscript:

- Marcelo Loor and Guy De Tré. *Enabling Augmented (Fuzzy) Computation in Social Media Mining.* Submitted for publication in *Fuzzy Sets and Systems.*

8.1 Introduction

A post on social media could reflect an opinion (or judgment) resulting from what one has learned or understood about a particular topic by experience. For instance, while one might consider a particular hotel to be a nice place to stay because it is peaceful, someone else might judge it to be a terrible hotel because it is far away from the main city attractions. Although it could be very subjective, information extracted from such posts may be considered to be helpful. If someone is choosing a hotel to stay in a not-yet-visited city, he/she might browse posts of his/her friends, colleagues or relatives who have already visited that city to know about their individual experiences. Nevertheless, due to large amounts of subjective data provided by heterogeneous sources, some (information) receivers might consider the resulting information to be out of context according to their individual understandings about the topic under discussion. While a quick summary might state: "seven out of ten people consider a particular hotel to be a nice place to stay," a more detailed summary might say: "six out of ten people consider the hotel to be a nice place to stay because it is far away of the noisy city attractions." If someone understood as a nice place to stay a "place that is peaceful," he/she will consider the quick summary to be helpful; otherwise he/she might consider it to be out of context (or even slightly misleading). A challenge that arises from this situation is *how to detect and manage automatically any difference in understanding of a topic (informally) discussed (or commented) through messages posted by persons considering a different context.*

In Chapter 6, we studied a closely related problem. Therein the challenge was to detect any difference in understanding of the topic *behind an evaluation request* in which the evaluations (or answers) could be provided by heterogeneous (human) sources. To address that challenge, a generalization of an *appraisal degree* within the framework of *intuitionistic fuzzy sets* [1, 2] was proposed. It was shown that such a generalization, named *augmented appraisal degree*, along with several operators and functions can deal with (similarity) comparisons of such experience-based (fuzzy) evaluations. Hence, one might expect that those *augmented computational tools* could help to address the problem stated above. However, since posts on social media usually do not result from an evaluation request, a specific problem to solve is *how to extract experience-based evaluations regarding a topic from posts that are not necessarily related to that topic.*

Aiming to address that specific problem, we propose in this chapter a novel computational intelligence method whereby posts on social media are digested to obtain *augmented (Atanassov) intuitionistic fuzzy sets* (AAIFSs) [3], each representing a collection of experience-based evaluations (XBEs) given by a specific person regarding a topic under analysis. To digest the posts, the proposed method involves steps that mimic the way in which a person performs an evaluation based on what he/she learned about the topic under analysis by experience.

As depicted in Figure 8.1, the *post-digest method* uses a collection $M_{@P}$

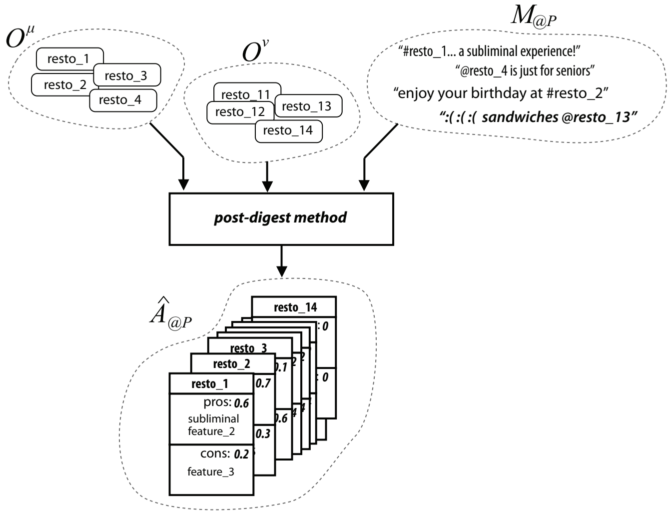


Figure 8.1: A general view of the proposed post-digest method.

together with the collections O^μ and O^ν as inputs to construct an AAIFS $\hat{A}_{@P}$. While the collection $M_{@P}$ consists of messages posted by a person P, the collections O^μ and O^ν contain objects that, according to an information seeker, say S, satisfy and dissatisfy respectively the criterion "be compatible with the way in which A is perceived," where A represents an idea or a mental picture (i.e., a concept) related to the topic under analysis – e.g., if the topic under analysis is about *nice places to stay*, then A will represent a mental picture of a *nice place to stay* and, thus, O^μ and O^ν will contain *places* that satisfy and dissatisfy respectively the criterion "be compatible with the way in which a *nice place to stay* is perceived," according to S. Regarding the AAIFS $\hat{A}_{@P}$, it characterizes the evaluations that result after learning from the messages posted by P what would be his/her understanding (or knowledge) about A to make the constituents of $O^\mu \cup O^\nu$ objects that satisfy or dissatisfy the aforementioned criterion.

An important aspect is that the AAIFSs resulting after digesting the messages posted by several users lend themselves to *augmented (fuzzy) computation*, i.e., those AAIFSs can be used in a process in which not only the judgments but also their contexts are taken into account for computation – herein, by *'context of a judgment'* we mean the conditions that arise when the message is posted, which mainly depend on the personal experience of the publisher. Thus, our method can be applied to build a kind of database containing AAIFSs characterizing XBEs given by a social media user with whom an information seeker shares a similar understanding about a given topic. A practical motivation here is that, if such a database is available for an information seeker, he/she can use it as a reliable source to obtain information about objects that he/she might not have experienced yet.

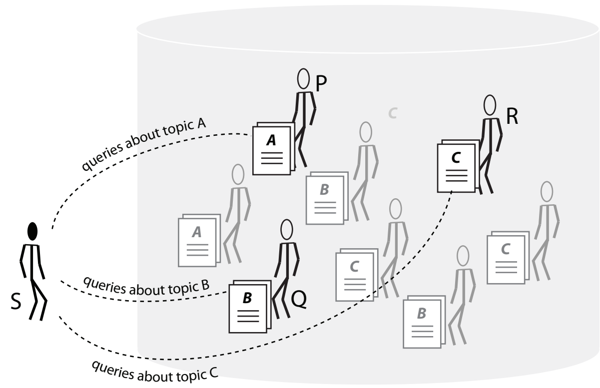


Figure 8.2: Practical motivation for this work.

For instance, Figure 8.2 depicts a situation where a database has been built based on what an information seeker S understands about topics A, B and C. Assuming in this case that S shares a quite similar understanding about topic A with a social media user P, S can find the data digested from the messages posted by P to be more reliable when looking for information about objects that are compatible with the way in which he/she perceives A. Likewise, if S shares a very similar understanding about topic B and C with social media users Q and R respectively, S can consider the data digested from the messages posted by Q and R to be more reliable to get information about objects that are compatible with the way to which he/she perceives B and C respectively.

Since such a database can be used to discover patterns or summarize content related to the different understandings of a concept that social media users may have, the proposed method could be implemented together with techniques used in opinion mining and information fusion [4] to produce even more personalized summaries according to the individual preferences of an information seeker. For instance, our method could be applied when someone would like to read a summary of recommended hotels that is obtained after processing only the posts published by persons having an understanding about "a nice place to stay" that is similar to his/hers.

Another important and interesting aspect of the post-digest method is that it allows a person to assess the quality of an information source without given details that might compromise his/her privacy. Hence, this method could be applied in situations where someone needs some privacy when looking for pertinent information given by a reliable source. For instance, if an information seeker, say Rod, is interested in a *nice place to stay*, he can put into O^{μ} and O^{ν} the names of some hotels that, according to his experience, are compatible and noncompatible respectively with the way he perceives *nice places to stay.* After using our method to digest the messages posted by him, as well as the messages posted by, e.g., Pia, Rod will obtain the AAIFSs $\hat{A}_{@Rod}$ and $\hat{A}_{@Pia}$. Each element in $\hat{A}_{@Rod}$ and $\hat{A}_{@Pia}$ represents a hotel with its augmented appraisal degree denoting the level to which the hotel is considered to be a nice hotel to stay and some hints on why this is the case. Thus, by performing an augmented comparison between these AAIFSs, Rod can measure the degree of similarity between his understanding and Pia's about (the topic) *"nice places to stay."* This measure can be used by Rod as an indicator of how reliable Pia is, acting as an information source for this topic without having to reveal information about his vacation plans.

To present the proposed *post-digest method*, this chapter has been structured as follows. The key ideas and definitions regarding AAIFSs are recalled in the next section. The internal structure and components of the post-digest method are described in Section 8.3. After that, in Section 8.4 we present an illustrative example that shows how the proposed method works. Before concluding with some suggestions on future research directions, we present some related work in Section 8.5.

8.2 Preliminaries

As was mentioned in the introduction, a post on social media can convey what someone thinks about a particular topic or concept, i.e., it can express an individual opinion or evaluation about a topic. Since such an evaluation can be influenced by a personal experience or knowledge, it could be fairly subjective. Moreover, it can be expressed in such a way that it is neither constrained to a *full agreement*, nor to a *full disagreement*, but can take any of the values in between to denote a partial agreement.

Aiming to make such an *experience-based evaluation* (XBE) available for computation, in Chapter 6 we proposed to model it by an *augmented appraisal degree* (AAD). Along with the definition of an AAD, the concept of *augmented (Atanassov) intuitionistic fuzzy sets* (AAIFS), as well as some computational tools were introduced. In this section, we briefly recall how one can use such an augmented framework to handle XBEs.

8.2.1 Augmented Appraisal Degrees

Consider that someone has posted the following message: "Hotel ABC is a fairly nice place to stay because it is near the beach." This message can be seen as an XBE in which an object, namely *'Hotel ABC'*, *fairly* satisfies the criterion 'be compatible with the way in which a *nice place to stay* is perceived' because this hotel is *near the beach.* The idea behind an AAD is that, instead of characterizing only the level to which this hotel satisfies the criterion, it can additionally include (some of) the reasons that suggest (or explain) *why* the criterion is fulfilled by the hotel.

According to Definition 6.1, the judgment in the above example can be characterized as an AAD, say $\hat{\ell}_{\mathcal{C}@P}(x)$, such that $\hat{\ell}_{\mathcal{C}@P}(x) = \langle \textit{fairly}, \{\textit{'located near the beach'}\}\rangle$, $x =$ *'Hotel ABC'*, $\mathcal{C} =$ 'be compatible with the way in which a *nice place to stay* is perceived', and P represents the person who posted the message.

To illustrate an advantage of using AADs to characterize XBEs, let us consider that another person, say Q, has posted this message: "Hotel ABC is a fairly nice place to stay because it is far from the noisy main city attractions." In this case, the judgment can be characterized as another AAD, say $\hat{\ell}_{\mathcal{C}@Q}(x)$, such that $\hat{\ell}_{\mathcal{C}@Q}(x) = \langle$*fairly*, $\{$*'located far from the noisy main city attractions'*$\}\rangle$. Notice that, although P and Q agree with the level to which x = *'Hotel ABC'* satisfies $\mathcal{C}$, the features considered by them to appraise x are different. If only the levels are considered, a similarity comparison between those judgments will result in a *pseudo-matching*, i.e., the judgments would match even though their contexts are different (see Chapter 3). Being able to avoid this kind of situations is considered to be an advantage of characterizing XBEs by AADs.

Since a criterion $\mathcal{C}$ is commonly related to the way to which a concept, say A, is perceived, for simplicity we shall use $\hat{\mu}_A(x)$ instead of $\hat{\ell}_{\mathcal{C}}(x)$ to denote an AAD resulting from the evaluation of (the proposition) 'x satisfies $\mathcal{C}$', as well as $\hat{\nu}_A(x)$ to denote an AAD resulting from the evaluation of 'x dissatisfies $\mathcal{C}$'. For instance, if A represents '*a nice place to stay,*' $\hat{\ell}_{\mathcal{C}@P}(x)$ and $\hat{\ell}_{\mathcal{C}@Q}(x)$ will be denoted by $\hat{\mu}_{A@P}(x)$ and $\hat{\mu}_{A@Q}(x)$ respectively in the above examples.

8.2.1.1 Comparing Two Augmented Appraisal Degrees

As suggested above, comparing two AADs involves both a comparison of their appraisal levels and a comparison of the collections of features that hint about those levels. To deal with the latter comparison, a *connotation alikeness factor* (CAF) has been proposed in Section 6.4.1 as an indicator of how similar the contexts of two AADs are – therein, the *context* of an AAD is shaped by the collection of features that hint the reasons of the appraisal.

As an example of using a CAF, let us consider that the message posted by a person R about *Hotel ABC* has been characterized as the AAD $\hat{\mu}_{A@R}(x) = \langle$*completely*, $\{$*'located near the beach', 'located near the big stores'*$\}\rangle$, where x = *'Hotel ABC'* and A denotes '*a nice place to stay.*' In this case, the CAFs $\Delta_{\mu_A:P,R@P}$ and $\Delta_{\mu_A:P,R@R}$ can be used to indicate, according to the perspectives of P and R respectively, how similar the collections of features in $\hat{\mu}_{A@P}(x)$ and $\hat{\mu}_{A@R}(x)$ are. Since the collections in $\hat{\mu}_{A@P}(x)$ and $\hat{\mu}_{A@R}(x)$ are $F_{\mu_A@P} = \{$*'located near the beach'*$\}$ and $F_{\mu_A@R} = \{$*'located near the beach', 'located near the big stores'*$\}$ in that order, these CAFs could be established by the person who performs the comparison as $\Delta_{\mu_A:P,R@P} = 1$ and $\Delta_{\mu_A:P,R@R} = 0.5$ respectively. Notice here that $\Delta_{\mu_A:P,R@P}$ is not equal to $\Delta_{\mu_A:P,R@R}$, i.e., the CAFs depend on which perspective is chosen as a point of reference, i.e., the perspective of P or R.

The comparison of the appraisal levels also depends on the point of view that is used as a reference. Thus, it is needed to determine how an AAD looks like when it is seen from a particular perspective. For that purpose, the '*as seen from*' operator $\lfloor\cdot\rfloor_{@Q}$ has been proposed in Definition 6.3. The idea behind this equation is that, when an AAD $\hat{\ell}_{\mathcal{C}@P}$ is seen from the perspective of Q, the collection of features detected by Q should be taken into account for the computation of the appraisal. Thus, $\lfloor \ell_{\mathcal{C}@P} \rfloor_{@Q}$ will not only depend on the value of $\ell_{\mathcal{C}@P}$ but also on how similar $F_{\mathcal{C}@P}$ and $F_{\mathcal{C}@Q}$ are. So, $\lfloor \ell_{\mathcal{C}@P} \rfloor_{@Q}$ will also depend on $\Delta_{\mathcal{C}:P,Q@Q}$, i.e., $\lfloor \hat{\ell}_{\mathcal{C}@P} \rfloor_{@Q} = \langle \ell_{\mathcal{C}@P} \cdot \Delta_{\mathcal{C}:P,Q@Q}, F_{\mathcal{C}@Q} \rangle$.

After finding out how an AAD looks like when it is seen from a particular perspective, it can be compared with another AAD given from that perspective.

8.2.2 Augmented (Atanassov) Intuitionistic Fuzzy Sets

Messages posted on social media can simultaneously include judgments about positive and negative aspects detected in objects related to the topic under analysis – e.g., a message could state *"I like #HotelABC because it's near the beach... but the downside is that it's too noisy sometimes."* As has been explained in Section 6.4.3, such judgments can be characterized as AADs and, so, an *augmented (Atanassov) intuitionistic set* (AAIFS) can be used to handle them.

As a way of illustration, consider that A and $\mathcal{C}$ are kept the same as in the above examples (i.e., A denotes *'a nice place to stay'* and $\mathcal{C}$ represents the criterion *'be compatible compatible with the way in which A is perceived'*) and $X = \{$*'Hotel ABC'*, *'Hotel DEF'*$\}$ is a collection of places to evaluate. Consider also that a unit interval scale is used to indicate the appraisal levels, where 1 denotes the highest level and 0 corresponds to the lowest level. Consider finally that a person, say P, makes the following judgments:

- *'Hotel ABC'* satisfies $\mathcal{C}$ at 0.65 because it is located near the beach; however, due to it is too noisy sometimes, it also dissatisfies $\mathcal{C}$ at 0.15.
- *'Hotel DEF'* satisfies $\mathcal{C}$ at 0.6 because of the hospitality of its staff and it is located near the big stores; even so, since it is expensive, this hotel also dissatisfies $\mathcal{C}$ at 0.3.

These judgments can be characterized as an AAIFS, say $\hat{A}_{@P}$, such that

$$\begin{aligned}\hat{A}_{@P} = \{ \ & \big\langle \textit{'Hotel ABC'}, \langle 0.65, \{\textit{'located near the beach'}\}\rangle \ , \\ & \qquad\qquad\qquad \langle 0.15, \{\textit{'too noisy sometimes'}\}\rangle \big\rangle \ , \\ & \big\langle \textit{'Hotel DEF'}, \langle 0.6, \{\textit{'located near the big stores', 'hospitality'}\}\rangle \ , \\ & \qquad\qquad\qquad \langle 0.3, \{\textit{'expensive'}\}\rangle \big\rangle \ \} \ .\end{aligned}$$

Here, the hesitation from P to judge *'Hotel ABC'* as *'a nice place to stay'* could be

$$\hat{h}_{A@P}(\textit{'Hotel ABC'}) = \langle 0.2, \{\textit{'shuttle service'}\}\rangle,$$

while his/her hesitation judging *'Hotel DEF'* could be

$$\hat{h}_{A@P}(\textit{'Hotel DEF'}) = \langle 0.1, \{\textit{'valet parking service'}\}\rangle.$$

8.2.2.1 Comparing two Augmented (Atanassov) Intuitionistic Fuzzy Sets

Analogous to a comparison of AADs, a comparison of two AAIFSs involves both a comparison of the appraisal levels and a comparison of their corresponding collections of features. In this case, the latter comparison uses a CAF between the collections containing *all* the features recorded in each AAIFS. This means that, if $\mathbf{F}_{\mu_A@P}(X)$ and $\mathbf{F}_{\mu_A@Q}(X)$ contain the features recorded in the AAIFSs $\hat{A}_{@P}$ and $\hat{A}_{@Q}$ respectively for each object $x_i \in X$ – i.e., $\mathbf{F}_{\mu_A@P}(X) = F_{\mu_A@P}(x_1)\cup\cdots\cup F_{\mu_A@P}(x_n)$ and $\mathbf{F}_{\mu_A@Q}(X) = F_{\mu_A@Q}(x_1)\cup\cdots\cup F_{\mu_A@Q}(x_n)$, a CAF $\Delta_{\mu_A:P,Q@P}$ will indicate from the perspective of P how similar the collections $\mathbf{F}_{\mu_A@P}(X)$ and $\mathbf{F}_{\mu_A@Q}(X)$ are. Likewise, a CAF $\Delta_{\nu_A:P,Q@P}$ will indicate from the same perspective how similar $\mathbf{F}_{\nu_A@P}(X)$ and $\mathbf{F}_{\nu_A@Q}(\mathrm{X})$ are.

Regarding the comparison of the appraisal levels, an *ℓ-measure* $\mathrm{sim}^{\alpha}_{\ell@P}$ defined by (6.21) was proposed in Section 6.4.3.1 as an option to determine the *level* to which $\hat{A}_{@P}$ and $\hat{A}_{@Q}$ are similar as seen from the perspective of P.

Using $\Delta_{\mu_A:P,Q@P}, \Delta_{\nu_A:P,Q@P}$ and (6.21), one can denote the result of the comparison between $\hat{A}_{@P}$ and $\hat{A}_{@Q}$ by a triplet $\langle \mathrm{sim}^{\alpha}_{\ell@P}, \Delta_{\mu_A:P,Q@P}, \Delta_{\nu_A:P,Q@P}\rangle$ or by a fusion of its components such as given by the equation

$$sim(\hat{A}_{@P}, \hat{A}_{@Q}) = (\lambda_\mu \Delta_{\mu_A:P,Q@P} + \lambda_\nu \Delta_{\nu_A:P,Q@P})\,\mathrm{sim}^{\alpha}_{\ell@P}, \qquad (8.1)$$

where $\lambda_\mu, \lambda_\nu \in [0,1]$ and $\lambda_\mu + \lambda_\nu \leq 1$. It is worth mentioning that, if a comparison between AAIFSs characterizing XBEs is denoted by a triplet, one can notice how similar the contexts of those XBEs are. As will be shown in Section 8.4, this representation can be useful in comparisons where not only the level but also the contexts of such evaluations are needed. In the next section, we describe how to obtain such evaluations by using the *post-digest method*.

8.3 Post-Digest Method

As was pointed out in the introduction to this chapter, we are interested in extracting experience-based evaluations (XBEs) related to a given topic from posts on social media that are not necessarily related to the topic. For this purpose, we introduce the *post-digest method*, which digests messages posted by a person on social media to obtain a collection of XBEs as if they were given by this person having the topic (or concept) under consideration in mind.

8.3.1 Extracting Experience-Based Evaluations

Since a message posted by a person may include some of the features that influence his/her judgment on an object satisfying or not a criterion related to a given concept, we could use such features to learn what would be his/her understanding about this concept. After learning so, we could use this understanding (or knowledge) to evaluate the level to which other objects are members or not of the concept according to this person's perspective – i.e., we could obtain XBEs of those objects as if these XBEs had been given from

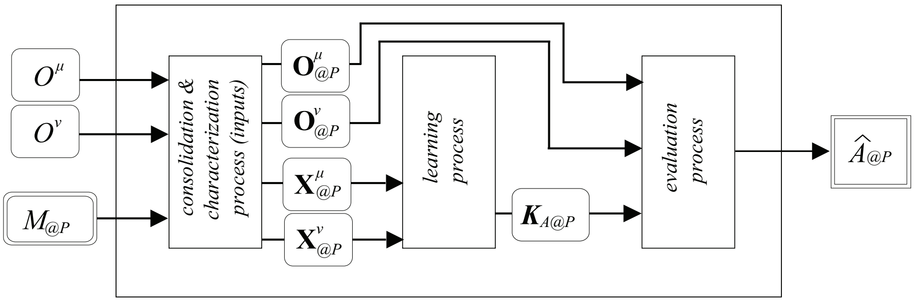


Figure 8.3: Internal structure of the post-digest method.

his/her perspective. We shall use this idea throughout the description of the post-digest method. Let us start this description with a formal statement of the problem regarding the extraction of XBEs from posts on social media:

> Consider the criterion *"be compatible with the way in which A is perceived,"* where A represents a particular concept. Consider also the collections O^μ and O^ν which include (identifiers of) objects that, according to an information seeker, say S, satisfy and dissatisfy respectively the aforementioned criterion. Consider finally a collection $M_{@P}$ consisting of messages posted by a person P. Let $\hat{A}_{@P}$ be an AAIFS that represents the XBEs of the objects in $O = O^\mu \cup O^\nu$ satisfying or dissatisfying the aforementioned criterion according to P. Under these considerations, find $\hat{A}_{@P}$ through the messages in $M_{@P}$ that are related to the objects in $O = O^\mu \cup O^\nu$.

To address this problem, we translate the ideas described in Chapter 2 into the design of the post-digest method depicted in Figure 8.3. As noticed, the post-digest method consists of three (sub) processes: (i) an input consolidation and characterization process, or ICC process for short; (ii) a learning process; and (iii) an evaluation process. In what follows, we describe each of them.

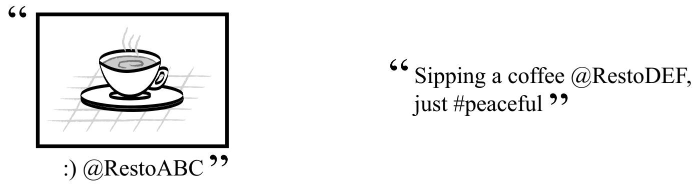
8.3.1.1 Input consolidation and characterization process

The purpose of the ICC process is to extract, consolidate and characterize the features related to the objects in $O^\mu \cup O^\nu$ that a message in $M_{@P}$ may have. To that end, the process uses (the identifiers of) the objects in $O^\mu \cup O^\nu$ and the messages in $M_{@P}$ as inputs to obtain the collections $\mathbf{X}^\mu_{@P}$, $\mathbf{X}^\nu_{@P}$, $\mathbf{O}^\mu_{@P}$ and $\mathbf{O}^\nu_{@P}$ as outputs (see Figure 8.3). The collections $\mathbf{X}^\mu_{@P}$ and $\mathbf{X}^\nu_{@P}$ will contain the *overall influence vectors* that, according to the feature-influence representational model (see Section 2.2.2), characterize the overall influence of the features extracted for each *pertinent message*, i.e., a message related to any object in $O^\mu \cup O^\nu$, that is found in $M_{@P}$. The collections $\mathbf{O}^\mu_{@P}$ and $\mathbf{O}^\nu_{@P}$ will contain the overall influence vectors that characterize consolidated objects having all the features extracted from all the pertinent messages. The steps of the ICC process are implemented in Algorithm 6.

Algorithm 6: ICC Process

Data: O^{μ}, O^{ν}, $M_{@P}$
Result: $\mathbf{O}^{\mu}_{@P}$, $\mathbf{O}^{\nu}_{@P}$, $\mathbf{X}^{\mu}_{@P}$, $\mathbf{X}^{\nu}_{@P}$
1 $\mathbf{O}^{\mu}_{@P}$,$\mathbf{O}^{\nu}_{@P}$, $\mathbf{X}^{\mu}_{@P}$, $\mathbf{X}^{\nu}_{@P} \leftarrow \{\}$
2 **foreach** $m_l \in M_{@P}$ **do**
3 $\quad F \leftarrow$ `extractFeatures(`m_l`)`
4 $\quad$ **foreach** $o_k \in (O^{\mu} \cup O^{\nu})$ **do**
5 $\quad\quad$ **if** `checkRelevance(`o_k,F`)` **then**
6 $\quad\quad\quad o_k.\mathcal{F}_k \leftarrow o_k.\mathcal{F}_k \cup F$ `/*` $o_k.\mathcal{F}_k$ `denotes the feature set of object` o_k `*/`
7 $\quad\quad\quad x_i.\mathcal{F}_i \leftarrow F$ `/* message` m_l `is represented by (features of) an object` x_i `*/`
8 $\quad\quad\quad$ **if** $o_k \in O^{\mu}$ **then** $X^{\mu} \leftarrow X^{\mu} \cup \{x_i\}$
9 $\quad\quad\quad$ **if** $o_k \in O^{\nu}$ **then** $X^{\nu} \leftarrow X^{\nu} \cup \{x_i\}$
10 $X \leftarrow X^{\mu} \cup X^{\nu}$
11 **foreach** $x_i \in X$ **do**
12 $\quad$ **foreach** $f_j \in x_i.\mathcal{F}_i$ **do**
13 $\quad\quad \beta_{i,j} \leftarrow$`getOverallWeight(`$f_j, x_i$`)`
14 $\quad\quad \mathbf{x}_i \leftarrow \mathbf{x}_i + \beta_{i,j}\hat{\mathbf{f}}_j$
15 $\quad$ **if** $x_i \in X^{\mu}$ **then** $\mathbf{X}^{\mu}_{@P} \leftarrow \mathbf{x}_i$
16 $\quad$ **if** $x_i \in X^{\nu}$ **then** $\mathbf{X}^{\nu}_{@P} \leftarrow \mathbf{x}_i$
17 **foreach** $o_k \in (O^{\mu} \cup O^{\nu})$ **do**
18 $\quad$ **foreach** $f_j \in o_k.\mathcal{F}_k$ **do**
19 $\quad\quad \beta_{k,j} \leftarrow$`getOverallWeight(`$f_j, o_k$`)`
20 $\quad\quad \mathbf{o}_k \leftarrow \mathbf{o}_k + \beta_{k,j}\hat{\mathbf{f}}_j$
21 $\quad$ **if** $o_k \in O^{\mu}$ **then** $\mathbf{O}^{\mu}_{@P} \leftarrow \mathbf{o}_k$
22 $\quad$ **if** $o_k \in O^{\nu}$ **then** $\mathbf{O}^{\nu}_{@P} \leftarrow \mathbf{o}_k$
23 **return** $\mathbf{O}^{\mu}_{@P}$, $\mathbf{O}^{\nu}_{@P}$, $\mathbf{X}^{\mu}_{@P}$,$\mathbf{X}^{\nu}_{@P}$

As can be noticed, the extraction of the features of each message in $M_{@P}$ is performed by the method *extractFeatures*, which returns the extracted features in a collection identified by F (see Line 3). Since the content of a message posted on social media can be fairly diverse, the method *extractFeatures* can involve several extraction techniques. For instance, to extract the emoticon, identifiers and the terms included in the posts depicted in Figures 8.4a and 8.4b, techniques connected with *natural language processing* [5] or *text categorization* [6, 7] such as *tokenization*, *stemming* or *stopword removal* [8, 9] can be used; analogously, to extract the features of the picture included in the post depicted in Figure 8.4a, techniques for object, scene or face recognition such as [10, 11, 12] can be applied. As will be shown in Section 8.4, the extraction of the features can also be performed by a large number of anonymous (human) contributors, using a *crowdsourcing* approach [13].



(a) (b)

Figure 8.4: Examples of posts on social media.

Previous to the consolidation of the extracted features, the relevance of a message is verified by the function *checkRelevance* (see Line 5). This function checks if one or more of the identifiers of an object o_k in $O^\mu \cup O^\nu$ are elements of the collection of extracted features F. If so, the message will be considered related to this object and, thus, the extracted features will be considered (and consolidated) as features of this object (see Line 6). The extracted features will also be used to represent the pertinent message under consideration, m_l, as an object x_i (see Line 7). After that, x_i is put into either X^μ or X^ν depending on which of the collections O^μ or O^ν the object o_k is part of (see Lines 8-9).

To characterize an object x_i in $X = X^\mu \cup X^\nu$ as a vector according to the feature-influence representational model, the overall weight $\beta_{i,j}$ of each feature f_j in $x_i.\mathcal{F}_i$ is computed by means of the method *getOverallWeight* (see Line 13). Next, the resulting overall influence vector $\mathbf{x}_i$ is computed (see Line 14). The method *getOverallWeight* can be implemented following information given by the information seeker or in accordance with the frequency of the extracted features. In the former case, an information seeker can for example consider that the overall weight of a *hashtag*, i.e., a word preceded by ‘#’, should be twice the overall weight of a (plain) word. In the latter case, the overall weight can be computed by, e.g., the equation

$$weight(f_j, x_i) = (1 + \ln(n(f_j, x_i))) \ln(|X|/n(f_j, X)) \tag{8.2}$$

proposed in [14] where $n(f_j, x_i)$ is the number of occurrences of f_j in x_i, $n(f_j, X)$ is the number of objects in X that contain f_j, and $|X|$ is the number of objects in X.

Similar steps are followed to characterize all the objects in $O^\mu \cup O^\nu$ as vectors (see Lines 17-22). Finally, all the vector collections are returned in Line 23.

As might be noticed, an important assumption of this method is that a pertinent message is posted by a person, say P, as an *“answer”* to an evaluation request, which is *“submitted”* by an information seeker. Hence, the content of this message could reflect (some of) the features that support the *“answer”* and, thus, these features can be used to learn something about P’s understanding of the concept (or topic) behind the evaluation request – this is the main reason why in Line 7 a pertinent message is represented by an object that will be used in the next learning process.

8.3.1.2 Learning process

The aim of the learning process is to obtain a model, say $K_{A@P}$, that represents the understanding of a person P about a concept A in which an information seeker S is interested. To do so, the learning method uses the collections $\mathbf{X}^{\mu}_{@P}$ and $\mathbf{X}^{\nu}_{@P}$, which contain the vectors that represent the pertinent messages posted by P, as *training collections* (see Figure 8.3).

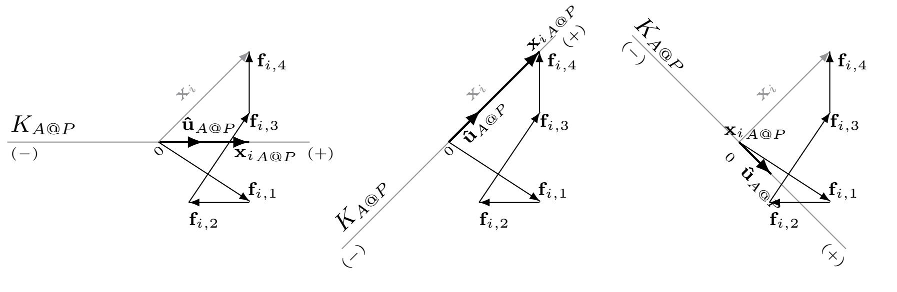
Following the idea presented in Chapter 2, the learning method tries to mimic a learning behavior in which one can learn about A by studying (the features of) the objects in $\mathbf{X}^{\mu}_{@P}$ and $\mathbf{X}^{\nu}_{@P}$ which satisfy (or dissatisfy) the criterion *"be compatible with the way in which A is perceived."* Thus, the main step of this method can be stated as a variant of the third step in the experience-based learning process presented in Section 2.2.3, i.e.

> Compute the constituents of $K_{A@P}$, i.e., $\hat{\mathbf{u}}_{A@P} = \omega_1\hat{\mathbf{f}}_1 + \cdots + \omega_m\hat{\mathbf{f}}_m$ and $t_{A@P}$, in such a way that (i) both the correspondence between each $\mathbf{x}_i \in \mathbf{X}^{\mu}_{@P}$ (or $\mathbf{x}_i \in \mathbf{X}^{\nu}_{@P}$) and the resulting specific influence of its features are preserved, and (ii) both the vector sum of the specific influences of the features of objects in $\mathbf{X}^{\mu}_{@P}$ and the vector sum of the specific influences of the features of objects in $\mathbf{X}^{\nu}_{@P}$ are maximized.

To recall how this step works, let us visualize it through the example presented in Figure 8.5, in which the following has been depicted:

- a line $K_{A@P}$ that represents a particular understanding of A, which is characterized by a directional vector $\hat{\mathbf{u}}_{A@P}$ and a threshold $t_{A@P}$ – for readability, the threshold $t_{A@P}$ has not been depicted;
- the vector $\mathbf{x}_i = \mathbf{f}_{i,1} + \mathbf{f}_{i,2} + \mathbf{f}_{i,3} + \mathbf{f}_{i,4}$, which represents the *resulting overall influence* of the features f_1, f_2, f_3 and f_4 of an object x_i when the proposition 'x_i satisfies the criterion *"be compatible with the way in which A is perceived"*' is appraised – recall from the *feature-influence* representational model described in Section 2.2.2 that, e.g., $\mathbf{f}_{i,2} = \beta_{i,2}\hat{\mathbf{f}}_2$ represents the *overall influence* of a feature f_2 when the fulfillment of the criterion is appraised on x_i, and $\beta_{i,2}$ denotes the *overall weight* (or importance) of f_2 among the features in x_i; and
- the vector $\mathbf{x}_{i\,A@P} = \mathbf{f}_{i,1\,A@P} + \mathbf{f}_{i,2\,A@P} + \mathbf{f}_{i,3\,A@P} + \mathbf{f}_{i,4\,A@P}$ which represents the *resulting specific influence* of the features of x_i when the criterion *"be compatible with the way in which A is perceived"*' is appraised on x_i – recall from the *feature-influence* representational model that, e.g., $\mathbf{f}_{i,2\,A@P} = \beta_{i,2\,A@P}\hat{\mathbf{u}}_{A@P}$ represents the *specific influence* of f_2 on x_i when the criterion is appraised on x_i, and $\beta_{i,2\,A@P}$ denotes the *specific weight* of f_2 during this appraisal. For readability, the vectors $\mathbf{f}_{i,1\,A@P}$, $\mathbf{f}_{i,2\,A@P}$, $\mathbf{f}_{i,3\,A@P}$ and $\mathbf{f}_{i,4\,A@P}$ have not been depicted.

During a learning process, $\hat{\mathbf{u}}_{A@P}$ and $t_{A@P}$ could be varied in order to increase (or decrease) the resulting specific influence of the features f_1, f_2, f_3 and f_4 when judging the criterion. For instance, turning $\hat{\mathbf{u}}_{A@P}$ counterclockwise, as



(a) (b) (c)

Figure 8.5: Varying the resulting specific influence of the features f_1, f_2, f_3 and f_4 of an object x_i when the fulfillment of the criterion “be compatible with the way in which A is perceived” is appraised on x_i.

shown in Figure 8.5b, yields an increment of the resulting specific influence of these features, i.e., the magnitude of $\mathbf{x}_{i\,A@P}$ increases. In contrast, turning $\hat{\mathbf{u}}_{A@P}$ clockwise, as shown in Figure 8.5c, makes that the resulting specific influence of these features disappears, i.e., the magnitude of $\mathbf{x}_{i\,A@P}$ becomes 0. The idea behind the learning process is to find potential *suitable* directional vectors and threshold points for a given concept, where ‘*suitable*’ means that the resulting specific influence of the features of each x_i in a training collection must correspond to the appraisal given for x_i. Thus, e.g., if x_i is an object that fulfills the criterion, i.e., $\mathbf{x}_i \in \mathbf{X}^{\mu}_{@P}$, then the resulting specific influence of its features must be in favor of the fulfillment of the criterion in such a way that the threshold $t_{A@P}$ is exceeded. After the potential suitable directional vectors and threshold points have been found, we must select the *optimal couple*, i.e., the couple $\langle \hat{\mathbf{u}}_{A@P}, t_{A@P} \rangle$ that maximizes both the aggregate of the specific influences of the features in favor of the criterion and the aggregate of the specific influences of the features in opposition to this criterion.

To find the optimal couple $\langle \hat{\mathbf{u}}_{A@P}, t_{A@P} \rangle$, one can use the procedure described in Section 2.2.4.

8.3.1.3 Evaluation process

The purpose of the evaluation process is to determine the level to which each object in $\mathbf{O} = \mathbf{O}^{\mu}_{@P} \cup \mathbf{O}^{\nu}_{@P}$ satisfies or dissatisfies the criterion *“be compatible with the way in which A is perceived”* according to the acquired knowledge $\mathbf{K}_{A@P}$ (see Figure 8.3).

As was mentioned in Section 2.2.2, the level to which an object x_i satisfies (or dissatisfies) the criterion corresponds to the level to which the *resulting specific influence* of its features exceeds (or is below) the threshold t_A according to K_A. Thus, the evaluation process consists of the computation of $\mathbf{l}_{i\,A@P} = \mathbf{x}_{i\,A@P} - t_A \hat{\mathbf{u}}_{A@P}$ (see (2.1)) for each $\mathbf{x}_i \in \mathbf{O}^{\mu}_{@P} \cup \mathbf{O}^{\nu}_{@P}$, where $\mathbf{x}_{i\,A@P}$ is the *vector projection* of (the resulting-overall-influence vector) $\mathbf{x}_i = \beta_{i,1}\hat{\mathbf{f}}_1 + \cdots + \beta_{i,m}\hat{\mathbf{f}}_m$

on $\hat{\mathbf{u}}_A = \omega_1\hat{\mathbf{f}}_1 + \cdots + \omega_m\hat{\mathbf{f}}_m$, i.e., $\mathbf{x}_{i\,A@P} = (\mathbf{x}_i \cdot \hat{\mathbf{u}}_{A@P})\,\hat{\mathbf{u}}_{A@P}$. However, since an AAIFS is expected as result, it is needed to represent each evaluation as an AAIFS element. To do so, we use the procedure presented in Chapter 7 as follows:

Let $O = \mathbf{O}^{\mu}_{@P} \cup \mathbf{O}^{\nu}_{@P}$; let x_i be one of the objects in O; and let f_j be a particular feature in $\mathcal{F}$. To identify whether f_j has been focused on during the judgment of the criterion on x_i, we compute the *specific influence* of this feature, i.e., $\mathbf{f}_{i,j_{A@P}} = \beta_{i,j\,A@P}\hat{\mathbf{u}}_{A@P}$, where $\beta_{i,j\,A@P} = \beta_{i,j}\omega_j$: when f_j has not been focused on, f_j does not reflect any influence, i.e., $\beta_{i,j\,A@P} = 0$; when f_j has been focused on and it is in favor of the criterion, the direction of $\mathbf{f}_{i,j_{A@P}}$ is the same as the direction of $\hat{\mathbf{u}}_{A@P}$, i.e., $\beta_{i,j\,A@P} > 0$; and when f_j has been focused on but it is in opposition to the criterion, the direction of $\mathbf{f}_{i,j_{A@P}}$ is opposite to the direction of $\hat{\mathbf{u}}_{A@P}$, i.e., $\beta_{i,j\,A@P} < 0$. Finally, to make the resulting levels conform to $I = [0, 1]$, we use (7.4) and (7.5).

At this point, all the internal processes of the post-digest method have been described. In the next section, we will show how these processes work together to obtain an AAIFS that represents a collection of XBEs.

8.4 Illustrative example

In this section, we present an example in which music album reviews were digested to detect reviewers who share a similar understanding about *top-rank albums*. To do so, in the first part we describe how the proposed method was configured and used for digesting the reviews and obtaining AAIFSs that represent XBEs of music albums. Then, in the second part, we explain how those AAIFSs were used in augmented comparisons aiming to detect the reviewers who share a similar understanding of what a top-rank album is.

In this example, we made use of a dataset containing music album reviews posted on Amazon.com between 1998-04-28 and 2014-07-23, which is part of the Amazon reviews[1] compiled in [15]. Among others, a review within this dataset consists of an identifier of a reviewer, an identifier of an album, a text describing the reviewer's judgment about the album, and an integer value between 1 and 5 that represents the score of the album assigned by the reviewer, where 1 and 5 are the lowest and the highest scores respectively. Accordingly, the following nomenclature is used throughout the example: (a) the collection of all the music album reviews in the dataset is denoted by $\mathcal{M}$; (b) the collection of the reviewers who have posted a review in $\mathcal{M}$ is denoted by $\mathcal{R}$; (c) the collection of the albums reviewed by any of the reviewers in $\mathcal{R}$ is denoted by $\mathcal{O}$; and (d) finally, a mental picture of *top-rank albums* is denoted by A.

8.4.1 Digesting music album reviews

Consider two reviewers in $\mathcal{R}$, say S and P. Let one of them, say S, be a reference reviewer (or someone who acts as an information seeker). Now, consider that

[1] https://snap.stanford.edu/data/web-Amazon.html

O_S^μ and O_S^ν are two collections containing the albums that, according to S, respectively satisfy and dissatisfy the criterion “be compatible with the way in which a *top-rank album* is perceived.” Consider finally a collection in $\mathcal{M}$, say $M_{@P}$, consisting of the reviews posted by P. In this context, we describe next how we configured and used the proposed method to digest $M_{@P}$ and, thus, obtain an AAIFS, say $\hat{A}_{S@P}$, as a result[2] – here, $\hat{A}_{S@P}$ will include evaluations that result after learning what would be P's understanding of A to make the albums in $O_S = O_S^\mu \cup O_S^\nu$ objects that satisfy or dissatisfy the aforementioned criterion.

To begin with, we implemented and configured the methods *extractFeatures* and *getOverallWeight* (see Algorithm 6) as follows. For the method *extractFeatures*, two variants were implemented: one variant using techniques connected to *natural language processing* and the other variant using *crowdsourcing*. In the first, the text included in a review is first split into words (or *tokens*) using separators such as commas, semi-colons, colons and blank-spaces. After that, *stop words* (i.e, words such as *‘among’*, *‘where’*, *‘too’*, etc. [8]) are removed from the list of words. Then, each word in the reduced list is stemmed using the Porter algorithm [9]. Finally, the resulting stemmed words are considered to be the features of the album evaluated in that review. Figure 8.6a shows the features extracted from a review included in the dataset after applying this variant of the method *extractFeatures*.

Regarding the crowdsourcing variant, we asked workers in the *Crowdflower*[3] platform to extract the features of the albums. To do so, we built a task in which a worker is presented with a review and asked to perform three steps: (1) read the review, (2) type an album's feature that seems to be important to the reviewer and (3) repeat step 2 as long as important features are identified. After collecting the provided features, we performed a de-duplication process to remove features that are perceived as being similar – e.g., *‘fun’*, *‘fun album’* and *‘fun and party feel’* were considered to be similar. Figure 8.6b shows the features extracted by applying this variant of the method *extractFeatures*.

For the method *getOverallWeight*, we implemented only one version, which is based on (8.2). In this example, each x_i in (8.2) characterizes a *pertinent* album and each f_j represents one of its extracted features.

After setting up the methods *extractFeatures* and *getOverallWeight*, we prepared the inputs of the post-digest method, i.e., $M_{@P}$, O_S^μ and O_S^ν as described below. While all the reviews posted by P were included into $M_{@P}$, the identifiers of the 10 best-rated and the 10 worst-rated albums reviewed by S were put into O_S^μ and O_S^ν respectively. Here, we assumed that the 10 best-rated and the 10 worst-rated albums reviewed by S satisfy and dissatisfy respectively the criterion “be compatible with the way in which a *top-rank album* is perceived.” It is worth mentioning that, to reduce the processing cost (in particular the crowdsourcing cost), we took into account only reviewers who have evaluated

[2]Since the example includes multiple reviewers acting as information seekers, the subscript S was included into the notation to keep track of the reviewer acting as an information seeker in a particular instance of the digest process.

[3]https://www.crowdflower.com

"For the most part, The Cars were a singles band, but their debut stands as one of the best ever new wave albums. The hit songs were solid, including the amazing threesome, "Let the Good Times Roll," "My Best Friend's Girl" and "Just What I Needed" that kick off the album. But side two is were The Cars demonstrate their artistic complexity, particularly on the lengthy double track "Moving in Stereo/All Mixed Up." "You're All I've Got Tonight" and "Bye Bye Love" make for another couple of great singles as well. This album is so good, it could stand as a greatest hits package all on its own.

part, car, singl, band, debut, stand, on, best, new, wave, album, hit, song, solid, includ, amaz, threesom, good, time, roll, friend, s, girl, need, kick, side, two, demonstr, artist, complex, particularli, lengthi, doubl, track, move, stereo/al, mix, up, re, ve, tonight, bye, love, make, anoth, coupl, great, well, greatest, packag

(a)

- solid hit songs
- artistic complexity
- greatest hits included

(b)

Figure 8.6: *Extracting the features of a particular music album that may be contained in the text of a review* . While the list (a) shows the features extracted using techniques connected to natural language processing, the list (b) shows the features extracted using crowdsourcing.

at least 6 albums with score greater than 3 and at least 6 albums with score less than 3.

Once the inputs were prepared, both variants of the post-digest method were executed for each pair of reviewers $\langle S, P\rangle$, where S is the reviewer acting as information seeker. Hence, a pair $\langle \hat{A}_{S@S}, \hat{A}_{S@P}\rangle$ of AAIFSs was obtained for each variant of the post-digest method. In the next part, we describe how a comparison of the AAIFSs in each pair was performed.

8.4.2 Finding reviewers with similar understandings

Since the AAIFSs in a resulting pair $\langle \hat{A}_{S@S}, \hat{A}_{S@P}\rangle$ contain the features that reviewers S and P would have focused on while posting their reviews, $\hat{A}_{S@S}$ and $\hat{A}_{S@P}$ can be compared to determine the level to which S and P share a similar understanding of what a top-rank album is.

As was mentioned in Section 8.2.2.1, a CAF between two AAIFSs can be used as an indicator of the similarity between the contexts of XBEs given by two persons. In this example, such a context is shaped by the features that influence the judgment of a reviewer about a music album being or not a top-rank album. This means that in this example a CAF for each pair $\langle \hat{A}_{S@S}, \hat{A}_{S@P}\rangle$ can be computed to identify reviewers with similar understandings of top-rank albums. Hence, to compute an approximation of $\Delta_{\mu_A:S,P@S}$, we made use of the equation

$$\begin{aligned}\Delta_{\mu_A:S,P@S} = n(\mathbf{F}_{\mu_A@S} \cap \mathbf{F}_{\mu_A@P}) / \, (n(\mathbf{F}_{\mu_A@S} \cap \mathbf{F}_{\mu_A@P}) \\ +\lambda_1 n(\mathbf{F}_{\mu_A@S} - \mathbf{F}_{\mu_A@P}) + \lambda_2 n(\mathbf{F}_{\mu_A@P} - \mathbf{F}_{\mu_A@S}))\,, \qquad (8.3)\end{aligned}$$

where $n(\mathbf{F}_{\mu_A@S} \cap \mathbf{F}_{\mu_A@P})$ denotes the number of common features, $n(\mathbf{F}_{\mu_A@S} - \mathbf{F}_{\mu_A@P})$ denotes the number of features that belong exclusively to $\mathbf{F}_{\mu_A@S}$,

$n(\mathbf{F}_{\mu_A @P} - \mathbf{F}_{\mu_A @S})$ denotes the number of features that belong exclusively to $\mathbf{F}_{\mu_A @P}$, and $\lambda_1, \lambda_2 \in [0,1]$ are parameters that adjust the contribution of $n(\mathbf{F}_{\mu_A @S} - \mathbf{F}_{\mu_A @P})$ and $n(\mathbf{F}_{\mu_A @P} - \mathbf{F}_{\mu_A @s})$ respectively. Analogously, we made use of the equation

$$\begin{aligned}\Delta_{\nu_A:S,P@S} = {} & n(\mathbf{F}_{\nu_A @S} \cap \mathbf{F}_{\nu_A @P}) / \left(n(\mathbf{F}_{\nu_A @S} \cap \mathbf{F}_{\nu_A @P}) \right. \\ & \left. + \lambda_1 n(\mathbf{F}_{\nu_A @S} - \mathbf{F}_{\nu_A @P}) + \lambda_2 n(\mathbf{F}_{\nu_A @P} - \mathbf{F}_{\nu_A @S}) \right), \end{aligned} \tag{8.4}$$

which is similar to (8.3), to compute $\Delta_{\nu_A:S,P@S}$. It is worth mentioning that (8.3) is based on the *ratio model* presented in [16] for describing the perceived similarity between two objects.

		P									
		R_4	R_{35}	R_7	R_{38}	R_9	R_2	R_{26}	R_5	R_6	R_{36}
S	R_4	1	0	0.03	0	0	0	0	0.09	0.06	0.09
	R_{35}	0.08	1	0.04	0.08	0.08	0.08	0	0	0.04	**0.35**
	R_7	0	0	1	0	0	0.06	0	0.09	0	0
	R_{38}	0	0.08	0	1	0.04	0	0	0	0	0.04
	R_9	0	0	0.04	0	1	0	0	0	0	0
	R_2	0.05	0	0.05	0	0.13	1	0	0.03	0.03	0
	R_{26}	0	0	0	0.08	0.08	0.08	1	0	0.08	0.04
	R_5	0.11	0	0.05	0	0	0	0	1	0.05	0
	R_6	0.12	0.04	0.04	0	0	0.04	0.04	0.08	1	0.12
	R_{36}	0.06	**0.14**	0	0	0.03	0.20	0	0.06	0.06	1

(a) $\Delta_{\mu_A:S,P@S}$

		P									
		R_4	R_{35}	R_7	R_{38}	R_9	R_2	R_{26}	R_5	R_6	R_{36}
S	R_4	1	0	0.07	0	0	0.13	0	0.03	0.07	0.03
	R_{35}	0	1	0.06	0.03	0.06	0.18	0	0.06	0.06	**0.09**
	R_7	0	0	1	0.05	0	0.05	0	0.03	0.05	0
	R_{38}	0	0	0	1	0.06	0.12	0	0	0	0
	R_9	0	0	0.03	0.03	1	0	0	0	0	0
	R_2	0.03	0	0.03	0	0.05	1	0	0.05	0	0
	R_{26}	0	0	0	0	0	0	1	0	0	0
	R_5	0.08	0.03	0.03	0	0	0	0	1	0.06	0
	R_6	0.05	0	0.02	0	0.10	0.12	0	0.07	1	0.02
	R_{36}	0	**0.09**	0	0	0.03	0.06	0	0.03	0.06	1

(b) $\Delta_{\nu_A:S,P@S}$

Table 8.1: CAFs between AAIFSs with features extracted using techniques connected to natural language processing.

The CAFs with the highest computed values are listed in Tables 8.1 and 8.2: while Table 8.1 shows the CAFs between AAIFSs with features extracted using techniques connected to natural language processing, Table 8.2 shows the CAFs between AAIFSs having features extracted using crowdsourcing.

S \ P	R_4	R_{35}	R_7	R_{38}	R_9	R_2	R_{26}	R_5	R_6	R_{36}
R_4	1	0	0	0	0	0	0	0.06	0	0
R_{35}	0	1	0	0	0	0	0	0	0.02	**0.02**
R_7	0	0	1	0	0	0	0	0	0	0
R_{38}	0	0	0	1	0	0	0	0	0	0
R_9	0	0	0	0.02	1	0	0	0	0	0
R_2	0.02	0	0.02	0	0	1	0	0	0.02	0
R_{26}	0.02	0	0	0	0	0	1	0	0.02	0
R_5	0.04	0	0	0	0	0	0	1	0.02	0
R_6	0.05	0.03	0	0	0	0.03	0	0.05	1	0
R_{36}	0	**0.09**	0	0	0	0	0	0	0.03	1

(a) $\Delta_{\mu_A:S,P@S}$

S \ P	R_4	R_{35}	R_7	R_{38}	R_9	R_2	R_{26}	R_5	R_6	R_{36}
R_4	1	0	0	0	0	0	0	0.02	0	0
R_{35}	0	1	0	0	0	0	0	0	0	**0.09**
R_7	0	0	1	0	0	0	0	0	0	0
R_{38}	0	0	0	1	0	0	0	0	0	0
R_9	0	0	0	0	1	0	0	0	0	0
R_2	0.05	0	0	0	0	1	0	0	0	0
R_{26}	0	0	0	0	0	0	1	0	0	0
R_5	0.02	0	0	0	0	0	0	1	0	0
R_6	0.03	0	0	0	0	0	0	0	1	0
R_{36}	0.03	**0.13**	0	0	0	0	0	0	0	1

(b) $\Delta_{\nu_A:S,P@S}$

Table 8.2: CAFs between AAIFSs with features extracted using crowdsourcing techniques.

Using the computed CAFs, one can detect the reviewers that might have a similar understanding about *top-rank albums*. For instance, in Table 8.1 one can observe that CAFs between the AAIFSs resulting after digesting messages posted by R_{35} and R_{36} are $\Delta_{\mu_A:R_{35},R_{36}@R_{35}} = 0.35$ and $\Delta_{\nu_A:R_{35},R_{36}@R_{35}} = 0.09$ respectively. This suggests that, although in a small number, reviewer R_{36} might have focused on the features considered by reviewer R_{35} during the "assessment" of *top-rank albums*. Notice that the CAF $\Delta_{\mu_A:R_{35},R_{36}@R_{36}} = 0.14$ suggests that reviewer R_{35} might also have focused on the features considered by reviewer R_{36} during the "assessment" – recall that we consider the reviews as if they were given as answers to a formal evaluation request about *top-rank albums*.

Regarding the method used to extract the features, one can notice that the CAFs between AAIFSs with features extracted using crowdsourcing have lower values (see Table 8.2). Yet, it is possible to detect which reviewers might have a more similar understanding about *top-rank albums*. For instance, since $\Delta_{\mu_A:R_{35},R_{36}@R_{36}} = 0.09$ is greater than $\Delta_{\mu_A:R_6,R_{36}@R_{36}} = 0.03$, one can say that R_{36}'s understanding seems to be slightly more similar to R_{35}'s than R_6's.

8.5 Related Work

A study about the contributions of fuzzy set theory to machine learning and data mining is presented in [17]. Although that study has mainly focused on fuzzy analysis rather than fuzzy data, the author indicates that contributions such as *interpretability*, *representation of uncertainty* and *incorporation of knowledge* seem to be useful for fuzzy data processing. Such contributions are evident in our work when an augmented (fuzzy) comparison between two AAIFSs resulting after digesting social media posts is performed.

A work about fuzzy computation in data mining [18] shows how information granules can be represented as fuzzy sets and how those granules can be processed using the fuzzy set framework during a data mining process. In a similar way, contextual information granules could be characterized as AAIFSs and processed using the mechanisms included into the augmented framework.

Regarding the importance of enabling augmented computation, the survey conducted in [19] highlights the positive effects of including context within an information fusion process. Some of the surveyed works show how context in *soft data* (or human judgments) could improve the quality of fused information. Hence, we foresee a potential use of AAIFSs characterizing digested information within this related area. As an example, we found in [20] a more specific area in which our method could be applied. The authors in that work use data fusion and mining techniques to propose a *reputation generation* procedure by which an indicator of the reputation of a product (or entity) is computed. A step within that procedure is the *opinion filtering* process, in which opinions that are not related with the target product are filtered out. Our method could help in this step to filter out opinions given by people having contrasting understandings of the topic under study.

About the interpretability of the post-digest method, a study about the importance (for an end user) of the interpretability of a model, an algorithm and the output(s) used during a fuzzy data mining process is presented in [21]. Agreeing with that, we have proposed the post-digest method and emphasized the benefits of getting an AAIFS as a result.

8.6 Conclusions

In this chapter we have described a computational intelligence method whereby messages posted by a person on social media are digested to obtain an augmented (Atanassov) intuitionistic fuzzy set (AAIFS), which characterizes a collection of experience-based evaluations (XBEs) given by this person regarding a topic under analysis.

Since the resulting AAIFSs lend themselves to augmented computation (i.e., those AAIFSs can be used in a process in which not only the judgments but also their contexts are taken into account for computation) one can measure to which degree the contexts of the characterized evaluations are alike (Research Question $Q3$). Thus, one can obtain an indicator of the similarity between the understandings that two persons might have about a particular topic (Research

Question $Q4$).

Enabling such an augmented computation is a key aspect of the proposed post-digest method since it can be implemented together with methods used in opinion mining or information fusion to produce personalized summaries according to the understanding possessed by an information seeker about a particular topic (Research Question $Q5$).

Another important aspect of the post-digest method is that it allows a person to identify people with a similar understanding about a topic without revealing the topic. Hence, the method could be used to assess the usefulness of information resulting from social media content without given details that might compromise the privacy of an information seeker.

The applicability of the proposed method for detecting information sources that demonstrate a common understanding of a topic, has been illustrated in an example in which music album reviews were digested to detect reviewers having a similar understanding about top-rank albums. The implementation of feature extraction components of the method, using techniques connected to natural language processing and crowdsourcing has also been illustrated in the example.

References

[1] Krassimir T. Atanassov. *Intuitionistic fuzzy sets.* Fuzzy sets and Systems, 20(1):87–96, 1986.

[2] Krassimir T. Atanassov. *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[3] Marcelo Loor and Guy De Tré. *On the need for augmented appraisal degrees to handle experience-based evaluations.* Applied Soft Computing, 54(Supplement C):284 – 295, 2017.

[4] Jorge A. Balazs and Juan D. Velásquez. *Opinion Mining and Information Fusion: A survey.* Information Fusion, 27(Supplement C):95 – 110, 2016.

[5] Robert Dale, Hermann Moisl, and Harold Somers. *Handbook of natural language processing.* CRC Press, 2000.

[6] Yiming Yang and Jan O. Pedersen. *A comparative study on feature selection in text categorization.* In ICML, volume 97, pages 412–420, 1997.

[7] David D. Lewis, Yiming Yang, Tony G. Rose, and Fan Li. *RCV1: A New Benchmark Collection for Text Categorization Research.* J. Mach. Learn. Res., 5:361–397, December 2004.

[8] Bing Liu. *Web data mining: exploring hyperlinks, contents, and usage data.* Springer Science & Business Media, 2007.

[9] M.F. Porter. *An algorithm for suffix stripping.* Program, 40(3):211–218, 2006.

[10] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. *Pictorial structures for object recognition.* International Journal of Computer Vision, 61(1):55–79, 2005.

[11] Koen E. Van De Sande, Theo Gevers, and Cees G. Snoek. *Evaluating color descriptors for object and scene recognition.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(9):1582–1596, 2010.

[12] Wenyi Zhao, Rama Chellappa, P. Jonathon Phillips, and Azriel Rosenfeld. *Face recognition: A literature survey.* ACM computing surveys (CSUR), 35(4):399–458, 2003.

[13] Daren C. Brabham. *Crowdsourcing.* Mit Press, 2013.

[14] Chris Buckley, Gerard Salton, and James Allan. *The Effect of Adding Relevance Information in a Relevance Feedback Environment.* In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '94, pages 292–300, New York, NY, USA, 1994. Springer-Verlag New York, Inc.

[15] Julian McAuley and Jure Leskovec. *Hidden factors and hidden topics: understanding rating dimensions with review text.* In Proceedings of the 7th ACM conference on Recommender systems, pages 165–172. ACM, 2013.

[16] Amos Tversky. *Features of similarity.* Psychological review, 84(4):327, 1977.

[17] Eyke Hullermeier. *Fuzzy sets in machine learning and data mining.* Applied Soft Computing, 11(2):1493 – 1505, 2011.

[18] K. Hirota and Witold Pedrycz. *Fuzzy computing for data mining.* Proceedings of the IEEE, 87(9):1575–1600, Sep 1999.

[19] Lauro Snidaro, Jesús García, and James Llinas. *Context-based Information Fusion: A survey and discussion.* Information Fusion, 25:16 – 31, 2015.

[20] Zheng Yan, Xuyang Jing, and Witold Pedrycz. *Fusing and mining opinions for reputation generation.* Information Fusion, 36:172 – 184, 2017.

[21] Christophe Marsala and Bernadette Bouchon-Meunier. *Fuzzy data mining and management of interpretable and subjective information.* Fuzzy Sets and Systems, 281:252 – 259, 2015. Special Issue Celebrating the 50th Anniversary of Fuzzy Sets.

Chapter 9

Reaching Consensus on Collective Experience-Based Evaluations

Case Study

Abstract

In Chapter 6, an *augmented framework* for handling experience-based evaluations (XBEs) given by a heterogeneous group of people has been studied. In this chapter, such an augmented framework is applied to a novel consensus reaching process in which the options considered within a decision-making process are evaluated by a heterogeneous group of experts. In this consensus reaching process, a moderator provides the participants with a collection of attributes (or features), which were taken into account by some of those participants, but were unobserved by others. Using this collection, the moderator can require each expert to refocus his/her attention on previously unobserved features and, thus, review his/her evaluations to increase the level of consensus on the *collective* evaluations of the options. An example that illustrates the steps of this process is presented.

This chapter is an adapted version of the following publication:

- Marcelo Loor, Ana Tapia-Rosero and Guy De Tré. *Refocusing Attention on Unobserved Attributes to Reach Consensus in Decision Making Problems Involving a Heterogeneous Group of Experts. Proceedings of: EUSFLAT- 2017 – The 10th Conference of the European Society for Fuzzy Logic and Technology, September 11-15, 2017, Warsaw, Poland IWIFSGN'2017 – The Sixteenth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets, September 13-15, 2017, Warsaw, Poland, Volume 2*, pages 405–416, 2018.

9.1 Introduction

A consensus in decision-making can be reached using a process where the participants (or experts), under the supervision of a moderator, can reconsider their evaluations to be in agreement with the group and, thus, to come to a commonly acceptable solution for the decision making problem under consideration [1, 2, 3]. Such a process is usually carried out to reach a collective decision on the most suitable option among a predefined collection of available options. Those options can be described by a collection of attributes which can also be defined in advance – one can find in the literature several studies about multi-attribute group decision-making in which the collection of attributes is predefined [4, 5, 6]. Herein, we use the word *'attribute'* to denote a *feature* or *characteristic* inherent to any of the available options.

Describing the options with a predefined collection of attributes (or features) can be useful when the participants of a decision-making process have similar education or expertise. However, when there are participants (or experts) having a different opinion on how the attributes should be evaluated, the problem of finding a consensus arises. As an example of such a situation, one can think about a family having to make a decision on which cookies to prepare. While dad wants somewhat sweet cookies, mom prefers no sugar at all and their little boy prefers very sweet cookies.

Aiming to find a consensus, in this chapter we propose a novel consensus reaching process in which the options are evaluated according to a flexible collection of attributes. Such flexibility gives the experts the freedom to express their evaluations on the basis of their personal interests, experience or knowledge. Hence, an expert can mainly focus on what he/she considers to be relevant during the evaluation of the options.

An interesting aspect in our proposal is that attributes initially unobserved by some experts – but observed by others – can be made available to all the participants. Thus, when a consensus reaching process is iteratively performed, a moderator aiming to increase the level of consensus can suggest the participants to refocus their attention on the updated collection of attributes. In this regard, our proposal can be used in a decision-making process in which a heterogeneous group of participants have to come to an overall consensus on a common solution without being limited to focus their attention on a particular collection of attributes. This aspect differentiates our process from others in which a predefined set of attributes is used throughout the entire decision-making process. An advantage of this approach is that a more enriched and informed collective decision can be reached.

Another interesting aspect of our method is that it can be applied to decision problems in which social media users are asked to evaluate the options, i.e., decision problems in which a very diverse group of (anonymous) people participate. For instance, a government agency can ask people about governmental policies in order to make these policies more reachable and effective, while other organizations can require their customers to assess the suitability of a product from their own perspectives [7]. Such people involvements in decision-making are becoming a common practice these days.

In the next section we present some examples illustrating how the evaluations given by an expert can be represented by means of the augmented framework proposed in Chapter 6. A first contribution of this chapter is the statement of the *flexible attribute-set group decision-making* (FAST-GDM) problem in Section 9.3.1. A second contribution is the description of the proposed consensus reaching process in Section 9.3.2. An illustrative example is given in Section 9.3.3. Related work is discussed in Section 9.4. The conclusions are stated in Section 9.5.

9.2 Preliminaries

A consensus reaching process can be partially or fully automatized when a proper mathematical representation of the evaluations given by the participants is available. Hence, in this section we present some applied examples of two concepts introduced in Chapter 6: *augmented appraisal degrees* and *augmented (Atanassov) intuitionistic fuzzy sets*, which were proposed to represent experience-based evaluations in such a way that they are available for computation.

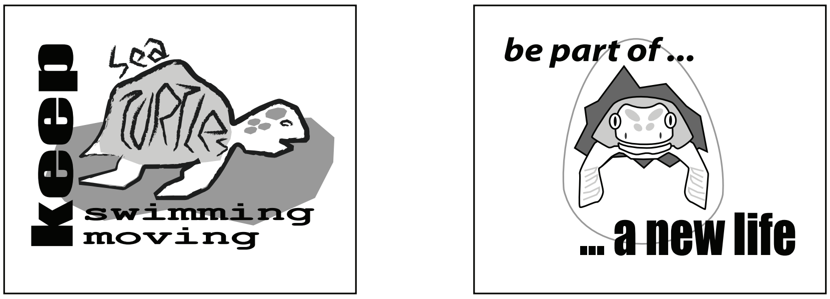
9.2.1 Augmented Appraisal Degrees in Decision-Making

When no constraint on the attributes (or features) of the options has been established, an evaluation might be accompanied by some clues about what have been focused on during the evaluation process – e.g., an expert might like to express that a particular option is quite suitable because of some of its features. This means that an evaluation can consist of the *level to which* an expert considers the option to be suitable for the problem under discussion, as well as a *collection of (some of) the reasons* justifying that judgment.

To characterize that kind of evaluations and make them available for computation, the idea of an *augmented appraisal degree*, AAD for short, was introduced in Section 6.4. That concept can be illustrated as follows:

Consider a collection $X = \{x_1, \cdots, x_n\}$ of discrete options for a particular problem. Consider also that a collection of *suitable options* for this problem in X is denoted by A, i.e., $A \subseteq X$. Consider finally an expert E who was asked to evaluate to which level an option $x_i \in X$ satisfies the proposition "p : x_i is a member of A." Let $\mathcal{F}_i$ be a collection of the *features* of x_i. In this context, an *augmented appraisal degree* of x_i, say $\hat{\mu}_{A@E}(x_i)$, is a pair $\langle \mu_{A@E}(x_i), F_{\mu_A@E}(x_i)\rangle$ that denotes the level $\mu_{A@E}(x_i)$ to which x_i satisfies p, as well as the collection of features $F_{\mu_A@E}(x_i)$ in $\mathcal{F}_i$ considered by E while appraising p.

As an example, let us consider that a marketing expert, say *Emma*, was asked to assess the level to which *'poster 1'* (see Figure 9.1a) is suitable for a campaign to protect sea turtles. Using a unit interval scale where 1 represents the highest level and 0 the lowest, *Emma* set 0.64 as her preferred level because of the *style of letters* and the *turtle's depiction.* In this case, after denoting



(a) Poster 1. (b) Poster 2.

Figure 9.1: Two options for a campaign to protect sea turtles.

the collection of *'suitable posters for a campaign to protect sea turtles'* by the letter A, one can characterize *Emma*'s appraisal with the AAD

$$\hat{\mu}_{A@Emma}(\text{'}poster\ 1\text{'}) = \langle 0.64, \{\text{'}style\ of\ letters\text{'},\ \text{'}turtle\text{'}s\ depiction\text{'}\}\rangle.$$

9.2.2 Augmented (Atanassov) Intuitionistic Fuzzy Sets in Decision-Making

In some situations, an expert might like to provide an evaluation denoting not only *how suitable* an option is, but also *how unsuitable* that option could be. For instance, the expert in the previous example, i.e, *Emma*, might like to express that the *flat face of the turtle* in *'poster 1'* makes it a bit unsuitable for the campaign. In other words, *Emma* might like to express that *'poster 1'* can and cannot be a member of the (collection of) *suitable options* at the same time.

To characterize that kind of judgments, the inclusion of AADs into the definition of an *intuitionistic fuzzy set* [8, 9], IFS for short, was proposed in Section 6.4.3. Such an augmented version of an IFS, named *augmented (Atanassov) intuitionistic fuzzy set*, AAIFS for short, can be described as follows:

Consider a discrete collection $X = \{x_1, \cdots, x_n\}$ of potential solutions, called *options*, for a particular problem. Consider also that $A \subseteq X$ is a collection of *suitable options* for this problem. Finally, consider an expert E who was asked to evaluate to which level an option $x_i \in X$ satisfies the propositions "$p : x_i$ is a member of A" and "$q : x_i$ is not a member of A." Let $\mathcal{F}_i$ be a collection of the *features* of x_i. Assume $\mathcal{I} = [0, 1]$ and $\mathcal{F} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_n$. Let $\hat{\mu}_{A@E}(x_i) = \langle \mu_{A@E}(x_i), F_{\mu_A@E}(x_i)\rangle$ and $\hat{\nu}_{A@E}(x_i) = \langle \nu_{A@E}(x_i), F_{\nu_A@E}(x_i)\rangle$ in $\langle \mathcal{I}, \mathcal{F}\rangle$ be two AADs respectively characterizing the appraisals of the propositions p and q given by expert E. In this context, an *augmented (Atanassov) intuitionistic fuzzy set* is a collection $\hat{A}_{@E}$ that describes the correspondence between each $x_i \in X$ and both $\hat{\mu}_{A@E}(x_i)$ and $\hat{\nu}_{A@E}(x_i)$ through the expression

$$\hat{A}_{@E} = \{\langle x_i, \hat{\mu}_{A@E}(x_i), \hat{\nu}_{A@E}(x_i)\rangle \mid (x_i \in X) \\ \wedge (0 \leq \mu_{A@E}(x_i) + \nu_{A@E}(x_i) \leq 1)\}\,. \tag{9.1}$$

As an example, let us consider the following judgments given by *Emma* about the suitability of both posters depicted in Figure 9.1 for the aforementioned campaign.

- *'Poster 1'* fulfills the proposition *"poster 1 is a suitable poster"* to an extent 0.64 because of the *style of letters* and the *turtle's depiction*; however, due to the *flat face of the turtle*, it also fulfills the proposition *"poster 1 is an unsuitable poster"* to an extent 0.11. The *presentation of the word "keep"* casts some doubt about the suitability of *"poster 1"* to an extent 0.25.
- *'Poster 2'* fulfills the proposition *"poster 2 is a suitable poster"* to an extent 0.58 because of the *cracked-egg concept* and the *slogan*; even so, since the *shape of the shell is confusing*, this poster also fulfills the proposition *"poster 2 is an unsuitable poster"* to an extent 0.23.

Assuming that the collection of posters (or options) is given by $X = \{$*'poster 1'*, *'poster 2'*$\}$ and the collection of suitable posters is represented with the letter A, one can represents *Emma*'s judgments with the AAIFS

$$\begin{aligned}\hat{A}_{@Emma} = \Big\{ &\Big\langle \textit{`poster 1'}, \langle 0.64, \{\textit{`style of letters'}, \textit{`turtle's depiction'}\}\rangle\ , \\ &\qquad \langle 0.11, \{\textit{`flat face of the turtle'}\}\rangle \Big\rangle\,, \\ &\Big\langle \textit{`poster 2'}, \langle 0.58, \{\textit{`cracked-egg concept'}, \textit{`slogan'}\}\rangle\ , \\ &\qquad \langle 0.23, \{\textit{`shape of the shell is confusing'}\}\rangle \Big\rangle \Big\}\,.\end{aligned}$$

Here, the hesitation of *Emma* to judge *'poster 1'* as a member of *'suitable options'* can be represented by

$$\hat{h}_{A@Emma}(\textit{`poster 1'}) = \langle 0.25, \{\textit{`presentation of the word ``keep'''}\}\rangle.$$

In the next section, the AAIFS concept will be used into a decision-making process in which a heterogeneous group of experts are given the facility to express their evaluations in a flexible way.

9.3 A Consensus Reaching Process over Alternatives having a Flexible Set of Attributes

As was mentioned in the introduction to this chapter, we aim to develop a novel consensus reaching process by which a heterogeneous group of experts can be asked to evaluate the options on the basis of a flexible collection of attributes.

Thus, in what follows we firstly show the formulation of a *flexible attribute-set group decision-making* problem. Secondly, we describe a novel consensus reaching process over options having a flexible collection of attributes. Thirdly, we present an illustrative example of the process.

9.3.1 Problem Statement

Using the AAIFS concept presented in Section 6.4.3, a *flexible attribute-set group decision-making* (FAST-GDM) problem can be formulated as follows:

Let $X = \{x_1, \cdots, x_n\}$ be a collection of discrete *options* for a problem under analysis. Let $A \subseteq X$ be a collection of *suitable options* for this problem. Let $E = \{E_1, \cdots, E_m\}$ be a collection representing a group of *experts* who were asked to evaluate to which level each option in X is member of A. Let

$$\begin{aligned} \hat{A}_{@E_j} = \big\{ &\langle x_i, \hat{\mu}_{A@E_j}(x_i), \hat{\nu}_{A@E_j}(x_i) \rangle \mid (x_i \in X) \\ &\wedge \big(0 \leq \mu_{A@E_j}(x_i) + \nu_{A@E_j}(x_i) \leq 1\big) \big\} \end{aligned} \tag{9.2}$$

be an AAIFS representing the evaluations given by expert E_j. Let

$$\begin{aligned} \hat{A} = \{ &\langle x_i, \hat{\mu}_{A}(x_i), \hat{\nu}_{A}(x_i) \rangle \mid (x_i \in X) \\ &\wedge (0 \leq \mu_{A}(x_i) + \nu_{A}(x_i) \leq 1) \} \end{aligned} \tag{9.3}$$

be an AAIFS representing the computed *collective evaluations* of the group of experts. Finally, let $\text{cix}(\hat{A}_{@E_j}, \hat{A})$ be a function, named *concordance index*, that computes the level of concordance between $\hat{A}_{@E_j}$ and $\hat{A}$ where a higher value denotes a higher concordance between them. In this context, the FAST-GDM problem boils down to *finding the most suitable option(s) with a general agreement among the experts*, that is, finding the most suitable option(s) in such a way that the average of all the *concordance indices*, i.e., $\frac{1}{m}\sum_{E_j \in E} \text{cix}(\hat{A}_{@E_j}, \hat{A})$, is maximized.

To solve a FAST-GDM problem, a moderator can adopt the following procedure:

1. Ask each expert to evaluate the options.
2. Determine the level of consensus on the evaluations.
3. If the level of consensus is not enough and asking the experts to perform a new round of evaluations is possible, give the experts a feedback and start all over again.
4. If the level of consensus is enough, choose the best option(s) based on the collective evaluations. Otherwise, notify that no consensus has been reached.

While the first three steps can usually constitute a *consensus reaching process* or CRP for short, the last step can be part of what is known as a *selection process*. In the next part, we describe in detail how the first three steps are implemented in the novel CRP, proposed in this chapter.

9.3.2 Consensus Reaching Process in FAST-GDM

To make it easier for a moderator to guide a consensus reaching process in a FAST-GDM problem, we propose the procedure implemented in Algorithm 7. The algorithm, named *flexible attribute-set consensus reaching algorithm* (FAST-CR), takes the following as inputs: a collection of experts (E), a collection of options (X), a consensus threshold (τ) and a number indicating how many iterations are possible while trying to reach a consensus (η). Using these inputs, the algorithm tries to obtain a collection of collective evaluations ($\hat{A}$) in such a way that the computed level of consensus is greater than or equal to the required consensus threshold. The algorithm returns both the collection of collective evaluations and a flag indicating whether or not a consensus has been reached after performing not more than τ iterations.

The FAST-CR algorithm is structured in four logical phases: *characterization, aggregation, quantification* and *feedback*. During the *characterization phase*, the evaluations given by each expert are characterized as an AAIFS (see Lines 7–8). This phase starts after the evaluations of all the experts have been received (see Line 6).

Through the *aggregation phase*, the AAIFSs that result from the characterization phase are aggregated to obtain the *collective* AAIFS $\hat{A}$ (see Lines 10–15). While the aggregation of the levels to which an option x_i is *suitable*, i.e., $\mu_A(x_i)$, is done in Line 11, the aggregation of the levels to which x_i is *unsuitable*, i.e., $\nu_A(x_i)$, is computed in Line 13. An average is used to aggregate the individual levels. The features focused on during the evaluations of the options are aggregated using the *union* operator: while the features favoring the suitability of x_i are aggregated in Line 12, the features disfavoring so are aggregated in Line 14 – here, each aggregation makes it possible to built a new collection with features that might be unobserved by some of the experts.

In the *quantification phase*, the collective level of consensus is computed (see Line 16). For that purpose, the average of the levels of concordance between the evaluations given by each expert and the computed collective evaluations is used. It should be noticed that, since several strategies can be used to compute a concordance index between two evaluation sets, an external call of the function *cix* is proposed in the algorithm. Hence, a moderator can, for example, choose a similarity measure designed to compare two experience-based evaluation sets [10] as a measure of the level of concordance.

During the *feedback phase*, the FAST-CR algorithm gives the experts feedback on their evaluations (see Lines 18–22). First, each expert is notified about the concordance index of his/her evaluations (see Line 19). Then, for each option x_i, a suggested action about how the expert should modify his/her evaluation is notified. In the case of the level of *suitability* of x_i (see Line 21), while $\mu_A(x_i) - \mu_{A@E_j}(x_i) > 0$ suggests that the expert should *increase* $\mu_{A@E_j}(x_i)$ with $|\mu_A(x_i) - \mu_{A@E_j}(x_i)|$ taking into account the attributes in $F_{\mu_A}(x_i)$, the expression $\mu_A(x_i) - \mu_{A@E_j}(x_i) < 0$ suggests that the expert should *decrease* $\mu_{A@E_j}(x_i)$ with that level. Analogously, when the level of *unsuitability* of x_i is considered (see Line 22), while $\nu_A(x_i) - \nu_{A@E_j}(x_i) > 0$ indicates that the expert should *increase* $\nu_{A@E_j}(x_i)$ with $|\nu_A(x_i) - \nu_{A@E_j}(x_i)|$ considering the

Algorithm 7: Flexible Attribute-Set Consensus Reaching Algorithm.

/* E: Experts; X: Options; τ: Consensus threshold; η: Maximum number of rounds */

Data: E, X, τ, η

/* *consensusReached*: true or false, $\hat{A}$: collective evaluations */

Result: *consensusReached*, $\hat{A}$

$\tau^* \leftarrow 0$ /* Current level of consensus */

$\eta^* \leftarrow 0$ /* Current number of rounds */

$n \leftarrow |X|$ /* Number of options */

$m \leftarrow |E|$ /* Number of experts */

repeat

  waitForAllEvaluations (E)

  /* Characterize the evaluations as AAIFSs */

  foreach $E_j \in E$ **do**

    $\hat{A}_{@E_j} \leftarrow$ transformEvaluationsToAAIFS (E_j)

  $\hat{A} \leftarrow \{\}$

  /* Aggregate the AAIFSs */

  foreach $x_i \in X$ **do**

    $\mu_A(x_i) \leftarrow \frac{1}{n} \sum_{E_j \in E} \mu_{A@E_j}(x_i)$

    $F_{\mu_A}(x_i) \leftarrow \bigcup_{E_j \in E} F_{\mu_{A@E_j}}(x_i)$

    $\nu_A(x_i) \leftarrow \frac{1}{n} \sum_{E_j \in E} \nu_{A@E_j}(x_i)$

    $F_{\nu_A}(x_i) \leftarrow \bigcup_{E_j \in E} F_{\nu_{A@E_j}}(x_i)$

    $\hat{A} \leftarrow \hat{A} \cup \{\langle x_i, \langle \mu_A(x_i), F_{\mu_A}(x_i)\rangle, \langle \nu_A(x_i), F_{\nu_A}(x_i)\rangle\rangle\}$

  /* Compute the current collective level of consensus */

  $\tau^* \leftarrow \frac{1}{m} \sum_{E_j \in E} \mathtt{cix}(\hat{A}_{@E_j}, \hat{A})$

  if $\tau^* < \tau$ **then**

    /* Give the experts feedback on their evaluations */

    foreach $E_j \in E$ **do**

      notify$(E_j$, 'Level of consensus:', $\mathtt{cix}(\hat{A}_{@E_j}, \hat{A}))$

      foreach $x_i \in X$ **do**

        /* The expression $\mu_A(x_i) - \mu_{A@E_j}(x_i) > 0$ suggests E_j to *increase* $\mu_{A@E_j}(x_i)$ with $|\mu_A(x_i) - \mu_{A@E_j}(x_i)|$ taking into account the features in $F_{\mu_A}(x_i)$ */

        /* The expression $\mu_A(x_i) - \mu_{A@E_j}(x_i) < 0$ suggests E_j to *decrease* $\mu_{A@E_j}(x_i)$ with the same value $|\mu_A(x_i) - \mu_{A@E_j}(x_i)|$. */

        notify$(E_j$, 'Suggested action:', $x_i, \mu_A(x_i) - \mu_{A@E_j}(x_i), F_{\mu_A}(x_i))$

        notify$(E_j$, 'Suggested action:', $x_i, \nu_A(x_i) - \nu_{A@E_j}(x_i), F_{\nu_A}(x_i))$

  $\eta^* \leftarrow \eta^* + 1$

until $(\tau^* \geq \tau)$ *or* $(\eta^* > \eta)$

consensusReached $\leftarrow (\tau^* \geq \tau)$

return *consensusReached*, $\hat{A}$

attributes in $F_{\nu_A}(x_i)$, the expression $\nu_A(x_i) - \nu_{A@E_j}(x_i) < 0$ indicates that the expert should *decrease* $\nu_{A@E_j}(x_i)$ with that level.

The steps in the above phases are repeated until the consensus threshold is reached or the permitted number of iterations is exceeded (see Line 24). In the next part, we present an example that illustrates how these steps can be followed for reaching consensus in a decision-making problem.

9.3.3 Illustrative Example

To demonstrate how the FAST-CR algorithm can help to reach consensus in a FAST-GDM problem, let us consider that, along with *Emma* two additional persons, say an animal rightist called *Fred* and a graphic designer called *Gia*, were asked to evaluate the level to which each poster in Figure 9.1 is suitable for a campaign to protect sea turtles (see Section 9.2).

In this example, we assume that the moderator has established $\tau = 0.8$ as the consensus threshold and $\eta = 5$ as the maximum number of rounds. We also assume that, for simplicity, the moderator has chosen a strategy in which only the levels of the evaluations are taken into account for the computation of a level of concordance. Hence, to make such computations, we will use a similarity measure, called *XVBr*, which is suitable for comparing XBEs characterized as IFS elements according to the empirical study presented in Chapter 4 – the interested reader is referred to Chapter 4 and Chapter 5 for a detailed description of this similarity measure and for its implementation respectively.

Table 9.1: Evaluation results denoting to which degree each poster in Figure 9.1 is considered to be suitable for a campaign to protect sea turtles (*Round 1*).

Poster	Suitability		Unsuitability	
	Level	Reason(s)	Level	Reason(s)
(a) *Evaluations by Emma.*				
1	0.64	style of letters, turtle's depiction	0.11	flat face of the turtle
2	0.58	cracked-egg concept, slogan	0.23	shape of the shell is confusing
(b) *Evaluations by Fred.*				
1	0.68	turtle's depiction, slogan	0.27	turtle seems to be floating
2	0.19	slogan	0.54	turtle's depiction
(c) *Evaluations by Gia.*				
1	0.82	place of the words, irregular strokes	0.18	turtle's shadow
2	0.23	style of letters	0.71	turtle's depiction
(d) *Collective Evaluations.*				
1	0.71	style of letters, turtle's depiction, slogan, place of the words, irregular strokes	0.19	flat face of the turtle, turtle seems to be floating, turtle's shadow
2	0.33	cracked-egg concept, slogan, style of letters	0.49	shape of the shell is confusing, turtle's depiction

To start with, after receiving the evaluations from *Emma*, *Fred* and *Gia*, listed in Table 9.1a, Table 9.1b and Table 9.1c respectively, we go through the *characterization phase* (Lines 7–8) to represent the evaluations as AAIFSs. Then, we complete the *aggregation phase* (Lines 10–15) to obtain the collection of collective evaluations $\hat{A}$ listed in Table 9.1d. After that, through the *quantification phase* (Line 16), we compute the individual concordance indices by means of the implementation of the similarity measure *XVBr*. The results of these computations are $\text{cix}(\hat{A}, \hat{A}_{@Emma}) = 0.44$, $\text{cix}(\hat{A}, \hat{A}_{@Fred}) = 0.87$ and $\text{cix}(\hat{A}, \hat{A}_{@Gia}) = 0.89$. Thus, the current collective concordance index in the first round is $\tau^* = 0.73$.

Because $\tau^* < \tau$ holds, no consensus has been reached, we start the *feedback phase* (Lines 18–22) by notifying *Emma* about her level of consensus, i.e., $\text{cix}(\hat{A}, \hat{A}_{@Emma}) = 0.44$, and the corresponding suggestions. For instance, since $\mu_A(\text{'poster 2'}) - \mu_{A@Emma}(\text{'poster 2'}) = -0.25 < 0$, Emmas is suggested to *decrease* with 0.25 the level to which she considers poster 2 to be *suitable* for the campaign and, since $\nu_A(\text{'poster 2'}) - \nu_{A@Emma}(\text{'poster 2'}) = 0.26 > 0$, she is suggested to *increase* 0.26 the level to which she considers poster 2 to be *unsuitable* for the campaign. She is also suggested to refocus her attention to the *style of the letters* in addition to the *cracked-egg concept* and the *slogan* of this poster for the new level of suitability, and the *turtle's depiction* in addition to the *confusing shape of the shell* for the new level of unsuitability – see the *reasons* for the *suitability* and *reasons* for the *unsuitability* of *Poster 2* that have been aggregated in the *collective evaluations* listed in Table 9.1d. In a similar way, *Fred* and *Gia* are given feedback on their evaluations. After completing the feedback phase, a new round of evaluations is started.

For illustration purposes, the evaluations obtained during the second round are listed in Table 9.2. Notice that, in this round, while *Emma* decreased the level of suitability of *'poster 2'* by 0.20 after considering the *style of letters*, *Gia* increased this level by 0.08 because of the *slogan*. Notice also that Fred decreased the level of unsuitability of *'poster 1'* by 0.06 because he now considers the *turtle's shadow* instead of its *apparent flotation*. After completing the second round, the individual concordance indices are $\text{cix}(\hat{A}, \hat{A}_{@Emma}) = 0.51$, $\text{cix}(\hat{A}, \hat{A}_{@Fred}) = 0.91$ and $\text{cix}(\hat{A}, \hat{A}_{@Gia}) = 0.92$. Hence the computed collective concordance level is $\tau^* = 0.78$.

Although no consensus has been reached in the first two rounds, and more rounds have to be performed, we can observe how the experts can express their evaluations according to what they consider to be relevant, and how the FAST-CR algorithm can deal with flexible collections of attributes (or features).

9.4 Related Work

This section presents some related work on consensus in decision-making problems in which a heterogeneous group of experts is involved. It also shows studies where extra flexibility on the options is provided.

Those studies are generally based on one of the following ideas or on a combination thereof: (i) schemes suggesting the more discordant experts to review their evaluations, and (ii) procedures adjusting the weights, representing the importance or relevance, associated to the experts. The latter refers to

Table 9.2: Evaluation results denoting to which degree each poster in Figure 9.1 is considered to be suitable for a campaign to protect sea turtles (*Round 2*).

Poster	Suitability		Unsuitability	
	Level	Reason(s)	Level	Reason(s)
(a) *Evaluations by Emma.*				
1	0.64	style of letters, turtle's depiction	0.11	flat face of the turtle
2	0.38	cracked-egg concept, slogan, *style of letters*	0.23	shape of the shell is confusing
(b) *Evaluations by Fred.*				
1	0.68	turtle's depiction, slogan	0.21	*turtle's shadow*
2	0.19	slogan	0.54	turtle's depiction
(c) *Evaluations by Gia.*				
1	0.82	place of the words, irregular strokes	0.18	turtle's shadow
2	0.23	style of letters, *slogan*	0.71	turtle's depiction
(d) *Collective Evaluations.*				
1	0.71	style of letters, turtle's depiction, slogan, place of the words, irregular strokes	0.17	flat face of the turtle, turtle's shadow
2	0.29	cracked-egg concept, slogan, style of letters	0.49	shape of the shell is confusing, turtle's depiction

studies where the weights are adjusted to obtain a weighted aggregation of individual evaluations and not necessarily by asking the experts to modify their evaluations.

In [11] a consensus framework that combines the aforementioned approaches has been presented. That framework includes a feedback mechanism where specific experts are suggested to renew their evaluations based on identification rules at different levels – attribute, alternative and global levels – and on a suggestion rule. It uses optimization algorithms to reassign the weights of experts on the attributes in order to reach the desired consensus.

Another work presents a consensus model that can be used when the level of knowledge among the experts is quite different [12]. The feedback mechanism of that model uses the weights associated to the experts during the aggregation of a collective evaluation, as well as throughout the feedback process. It is assumed that the highest weight is given to experts having a deeper knowledge about the problem, so that they need a smaller support than those with a lower weight. Thus, the experts are given advice according to their weight or level of importance, e.g., when the consensus level is low, more advice to revise their opinions is provided to the experts with a low importance level.

Another model that introduce some flexibility during the decision process has been presented in [13]. In this model, the experts could provide their evaluations with heterogeneous information, i.e., by means of numeric or linguistic evaluations over pairwise comparisons over the alternatives. An interesting aspect of the model is that it uses a mechanism that allows a moderator to insert good alternatives or to remove the unfeasible alternatives during the next discussion rounds.

In [14], it has been presented a more recent work in which a changeable collection of options is used. In that work it is considered that the collection of experts can also change, i.e., experts can abandon or enter the discussion. If an expert would like to participate the involved experts should decide on his/her participation by means of a majority approach, but if an expert would like to abandon the discussion the process continues without taking into account the evaluations provided by that expert.

Although several works providing flexibility over the options exist, to the best of our knowledge none of them uses options having a flexible set of attributes (or features).

9.5 Conclusions

In this chapter, we have proposed a novel consensus reaching process over options having a flexible collection of attributes (or features). This process, which has been implemented in a *flexible attribute-set consensus reaching algorithm* (FAST-CR) considering a flexible collection of features describing the context of the evaluations, is intended to be used in decision-making problems involving a group of participants or experts with different personal interests, experience or knowledge (Research Question Q5). This led us to the formulation of the *flexible attribute-set group decision-making problem* (FAST-GDM).

In contrast to approaches in which the participants are given a predefined set of attributes on which evaluation criteria are specified to evaluate the options, in our approach the participants can perform those evaluations based on the features they consider to be worthy of attention. This is useful in situations in which, to increase the level of concordance, a moderator would like to put under consideration some features that might be observed by some participants but unobserved by others.

To represent such flexible evaluations and make them available for computation, augmented (Atanassov) intuitionistic fuzzy sets (AAIFSs) are used in our novel solution for the FAST-GDM problem and, also, into the implementation of the current version of the FAST-CR algorithm. However, the use of other kind of representations is suggested and subject to further study.

References

[1] Denis Bouyssou, Didier Dubois, Henri Prade, and Marc Pirlot. *Decision Making Process: Concepts and Methods.* John Wiley & Sons, 2013.

[2] Dominika Gołuńska and Janusz Kacprzyk. *A Consensus Reaching Support System for Multi-criteria Decision Making Problems.* In Challenging Problems and Solutions in Intelligent Systems, pages 219–235. Springer, 2016.

[3] Witold Pedrycz, Petr Ekel, and Roberta Parreiras. *Fuzzy multicriteria decision-making: models, methods and applications.* John Wiley & Sons, 2011.

[4] Yucheng Dong, Jing Xiao, Hengjie Zhang, and Tao Wang. *Managing consensus and weights in iterative multiple-attribute group decision making.* Applied Soft Computing, 48:80 – 90, 2016.

[5] Deng-Feng Li, Zhi-Gang Huang, and Guo-Hong Chen. *A systematic approach to heterogeneous multiattribute group decision making.* Computers & Industrial Engineering, 59(4):561–572, nov 2010.

[6] Yang Liu, Zhi-Ping Fan, and Xiao Zhang. *A method for large group decision-making based on evaluation information provided by participators from multiple groups.* Information Fusion, 29:132–141, 2016.

[7] 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.

[8] Krassimir T. Atanassov. *Intuitionistic fuzzy sets.* Fuzzy sets and Systems, 20(1):87–96, 1986.

[9] Krassimir T. Atanassov. *On Intuitionistic Fuzzy Sets Theory*, volume 283 of *Studies in Fuzziness and Soft Computing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[10] Marcelo Loor and Guy De Tré. *In a Quest for Suitable Similarity Measures to Compare Experience-Based Evaluations.* In Juan Julián Merelo, Agostinho Rosa, José M. Cadenas, António Dourado Correia, Kurosh

Madani, António Ruano, and Joaquim Filipe, editors, Computational Intelligence: International Joint Conference, IJCCI 2015 Lisbon, Portugal, November 12-14, 2015, Revised Selected Papers, volume 669 of *Studies in Computational Intelligence*, pages 291–314. Springer International Publishing, 2017.

[11] Chao Fu, Michael Huhns, and Shanlin Yang. *A consensus framework for multiple attribute group decision analysis in an evidential reasoning context.* Information Fusion, 17:22–35, 2014.

[12] Ignacio J. Pérez, Francisco J. Cabrerizo, Sergio Alonso, and Enrique Herrera-Viedma. *A new consensus model for group decision making problems with non-homogeneous experts.* IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44(4):494–498, 2014.

[13] Ignacio J. Pérez, Francisco J. Cabrerizo, and Enrique Herrera-Viedma. *A mobile group decision making model for heterogeneous information and changeable decision contexts.* International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 19(supp01):33–52, 2011.

[14] Juan A. Morente-Molinera, Rami Al-Hmouz, Ali Morfeq, Abdullah Saeed Balamash, and Enrique Herrera-Viedma. *A Decision Support System for Decision Making in Changeable and Multi-Granular Fuzzy Linguistic Contexts.* Journal of Multiple-Valued Logic & Soft Computing, 26, 2016.

Chapter 10

Conclusions and Further Research

10.1 Conclusions

In Chapter 1, we identified three challenges that arise when dealing with experience-based evaluations (XBEs) coming from a large number of (potentially anonymous) respondents. While the first challenge is concerned with the *characterization of XBEs*, the second and the third challenges are related, respectively, to the *methods needed to process XBEs* and to the *quality of such XBEs*. The next research questions have been formulated according to these challenges:

Q1. How to characterize subjective, imprecise and potentially marked-by-hesitation XBEs in such a way that they are suitable for computation?

Q2. How to perform a reliable comparison between XBEs given by a heterogeneous group of respondents?

Q3. How to measure the perceived quality of XBEs according to a particular understanding?

Q4. How to identify XBEs given by (anonymous) respondents with whom a requester shares a similar understanding of the topic under analysis?

Q5. How to detect and manage automatically any difference in understanding of a concept behind an evaluation request, in which the answers could be given by respondents with different background?

In what follows, we draw our conclusions about how this dissertation contributes to address these challenges and to answer these research questions.

10.1.1 About the characterization of XBEs

Keeping in mind that studies in the area of computational intelligence aim to find not completely accurate solutions but achievable and robust ones, in Chapter 2 we have described our interpretation on how a person can experience a concept and how this experience is then reflected in his/her XBEs related to this concept. With this interpretation, we have illustrated how the context of an XBE, i.e., the conditions that arise when the evaluation is carried out, is mainly influenced by the experience (or knowledge) acquired by a person while learning the concept under evaluation.

An insight obtained from that interpretation is that "*comparing XBEs having hints about what aspects have influenced the evaluation can be more reliable than comparing XBEs without such hints.*" Hence, we have studied the implications of characterizing XBEs *with* and *without* such hints: while in the latter case an XBE is characterized as an element of an *Atanassov intuitionistic fuzzy set*, or IFS short (see Chapter 3), in the former case an XBE is represented as an element of an *augmented (Atanassov) intuitionistic fuzzy set*, AAIFS for short. Such an element on its turn is characterized by an *augmented appraisal degree*, or AAD for short, which is a novel generalization of a membership (or non-membership) degree (see Chapter 6).

It is worth mentioning that, since the area of computational intelligence is quite extensive and XBEs are inherently human statements, the aforementioned characterizations have been based on concepts connected with *fuzzy set theory*, which exploit the fact that (information extracted from) human statements can be imperfect, but valuable for people.

We have shown that, when XBEs are provided by a *homogeneous group* of respondents, i.e., when the providers of XBEs are people having a similar understanding of the concept under evaluation, those characterizations by elements of an IFS or AAIFS are suitable for representing subjective, imprecise and potentially marked-by-hesitation XBEs, i.e., in such a case both characterizations can be deemed to be an answer to Research Question Q1. However, when a representation of XBEs provided by a *heterogeneous group* is needed, characterizing XBEs as elements of an AAIFS has an advantage since the context of an XBE can be recorded. This advantage becomes obvious when, e.g., XBEs are compared to each other.

10.1.2 About methods and techniques needed to process XBEs

In this dissertation we have centered our attention on comparison procedures that can be used to estimate a perceived similarity among XBEs. We have done so because such comparison procedures are in most cases needed to classify, filter or arrange XBEs.

Given that a comparison between two XBEs depends on their characterization, we have studied comparison procedures within the IFS framework, as well as comparison procedures included in the novel AAIFS framework described in Chapter 6.

In the case of XBEs characterized as elements of an IFS, since the aspects that have influenced an XBE cannot be recorded in an IFS element, only the appraisal levels can be taken into account for comparisons between IFSs. Hence, in Chapter 3 we have proposed a novel approach to compare XBEs that are characterized as IFS elements. In this approach, a kind of footprint of the comparison, called *connotation-differential print* or CDP for short, is built using the XBEs of some *relevant* objects, which have features that make them good examples of compatibility (or incompatibility) with the concept under evaluation. The idea behind this approach is that *"a difference in understandings could be marked by a difference in one or more XBEs of relevant objects."* It is worth mentioning that such an idea has emerged from the interpretation described in Chapter 2 on how the knowledge (or experience) acquired by a respondent is reflected in his/her XBEs.

In Chapter 4, we have shown by means of an experimental study that a CDP can help to perform a simple but acceptable comparison between IFSs characterizing XBEs given by a heterogeneous group of respondents. In this study, some existing similarity measures used in the IFS framework, as well as some novel similarity measures that have been augmented with a CDP, were tested with IFSs characterizing simulated XBEs given by people who learned a concept under different scenarios.

The results of the experimental study suggest that the IFS framework enriched with such novel *augmented* similarity measures has a good potential to be an answer to Research Question Q2. Further research and more elaborated experiments are needed in this regard. Hence, in Chapter 5 we have proposed an open-source software package, named *IFSMetrics*, by which a researcher or practitioner can replicate, modify or extend our study with new data sets or novel similarity measures.

Regarding XBEs characterized as elements of an AAIFS, one can use the aspects recorded in two AADs to determine the level to which the characterized XBEs are similar to each other. The idea behind this approach is that, *if two persons have the same (or very similar) understanding of the concept under evaluation, they will focus on the same (or very similar) aspects of an object in order to make a judgment about it.* This means that a comparison of two XBEs could not only be affected by the magnitude of each appraisal, but also by the aspects that the providers of these XBEs have focused on according to their individual understandings of the concept under evaluation.

To deal with such comparisons, in Chapter 6 we have proposed a novel *augmented framework* for handling AADs. In this framework, several concepts, methods and operators have been defined. Among them, a *connotation alikeness factor* (CAF) has been proposed as an indicator of the level to which the aspects that have influenced two (collections of) XBEs are alike. Along with the CAF concept, *comparison operators* like '=' or (its fuzzy variant) *'approximately equal'* have been proposed to compare two AADs that respectively characterize two XBEs. To handle collections of XBEs, novel methods for computing the similarity between two AAIFSs have been proposed.

We have shown that, even if two collections of XBEs are given by persons having different understandings of the concept under evaluation, one can perform a meaningful (and more reliable) comparison between them by computing the similarity between their corresponding AAIFSs. Hence, the augmented framework constitutes a good option to address Research Question Q2.

10.1.3 About the quality of XBEs

In this dissertation, it has been considered that *high-quality XBEs* are XBEs that are fit for use by a requester. Thus, we have deemed *usefulness* and *usability* to be important aspects of the quality of XBEs: while *usefulness* is connected with the fact that an XBE can be used, *usability* is related to the level to which an XBE is fit for use.

Considering that the *relevance* of an XBE is strongly linked to the usability of an XBE (and, so, to its perceived quality), we have paid special attention to this characteristic. In this regard, we have considered that the relevance of the XBEs given by a respondent depends on how aligned his/her understanding of the concept under evaluation is in relation to the understanding possessed by a requester. Herein such XBEs are deemed to be a consequence of an evaluation request in which a requester asks a (usually large) group of (heterogeneous) respondents to evaluate a collection of objects.

As was mentioned in the previous section, both the proposed AAIFS framework and the regular IFS framework with *augmented* similarity measures can be used for measuring the level to which two persons share a similar understanding of a particular concept. Therefore, we can consider that both frameworks are viable options to measure the level to which one or more XBEs are relevant according to a particular understanding, i.e., both frameworks are options to answer Research Question Q3.

Since one can expect that XBEs given by respondents with whom a requester shares a similar understanding are more reliable for him/her than XBEs given by respondents having a different understanding, we have made use of the aforementioned frameworks to differentiate such XBEs, i.e., we have deemed those frameworks to be practical options to answer Research Question Q4.

In that regard, in Chapter 7 we have proposed a novel method, called the *k-well-(un)fitted-specimens method*, to detect whether the contexts of crowdsourced XBEs on social media content are in line with the context of the XBEs given by a requester. Handling *plain* XBEs, i.e., XBEs in which only appraisal levels are available, is an important aspect of this method because it can be applied in situations where hints on the assessments are not available. By means of an empirical study with simulated XBEs, we have provided adequate evidence of the effectiveness of the *k-well-(un)fitted-specimens method* for identifying and measuring potential differences in the understandings that a heterogeneous group of respondents may have about the concept under evaluation. Thereby, we have also provided a practical option to answer Research Question Q5.

In the same regard, in Chapter 8 we have proposed another novel method, called the *post-digest method*, by which *unrequested* opinions posted on social media are digested to obtain AAIFSs characterizing XBEs. Since such AAIFSs can be used for measuring the level to which the contexts of the characterized XBEs are alike, one can use the method to detect opinions (or messages) posted by people sharing a similar understanding about a given fact or topic. By doing so, one can extract more reliable information when processing such messages. We have illustrated the applicability of the post-digest method with an example in which music album reviews were digested to detect reviewers having a similar understanding of top-rank albums. Hence, we can consider that the post-digest method helps to answer Research Question Q4.

The *k-well-(un)fitted-specimens* and *post-digest* methods are somehow oriented to the *selection* of the best XBEs (and, so, the best sources), based on how relevant these XBEs are for a particular requester (or information seeker). However, in some cases the relevance of XBEs needs to be determined not by an individual, but by a group of persons. To illustrate such a case, we have presented a novel consensus reaching process, called the *flexible attribute-set consensus reaching (FAST-CR)* process, in Chapter 9.

This consensus reaching process is intended to be used in decision-making problems involving a heterogeneous group of participants or experts where the relevance of the XBEs will depend on whether or not a consensus among the experts is reached. We have shown how the FAST-CR process can handle XBEs given by respondents with different personal interests, experience or knowledge. Hence, it constitutes another practical option to answer Research Question Q5.

10.2 Further Research

In this dissertation we have made important contributions to handle XBEs provided by a heterogeneous group of respondents. However, we have identified several topics that need to be further investigated.

One of such topics is the *augmentation of other frameworks* that might be applicable for the characterization and handling of XBEs. For instance, it would be interesting to augment bipolar satisfaction degrees and elements of a Pythagorean fuzzy set, and design methods that compare them in accordance with a particular perspective. We anticipate that compelling semantic interpretations caused by, e.g., *overspecified* bipolar satisfaction degrees can be more adequately handled in this way.

Another such a topic is the *aggregation of XBEs.* Herewith, we can ask ourselves if and how XBEs with different contexts can be aggregated to produce, e.g., personalized summaries that are in accordance with the individual understanding of a given concept possessed by an information seeker. Someone may ask in this regard if combining two or more XBEs having different contexts should be deemed to be an *aggregation*, i.e., the result of combining two elements with common characteristics, or a *fusion*, i.e., the result of combining elements with different characteristics.

The study of new methods and strategies to estimate the level to which the contexts of XBEs are alike is also recommended as further research. In this regard, it would be interesting to study a strategy in which some of the hints of a judgment are explicitly considered to be more important than others.

The effect of characteristics like *believability* or *added-value* in the perceived quality of XBEs is also a suggested study. In this case, someone might ask if, e.g. during a decision-making process involving a heterogeneous group of participants, a large number of aspects justifying an XBE increases the perceived value of this XBE because taking into account more aspects might facilitate the consensus reaching process.

10.3 Overall Conclusion

10.3.1 Contributions per Chapter

The novel contributions of each chapter in this dissertation are summarized as follows.

- The three challenges that we identified while handling XBEs coming from a large number of (potentially anonymous) respondents have been introduced in Chapter 1. The purpose, the practical motivations and the scope of this dissertation have also been presented in Chapter 1.

- Insights about *what aspects might have an influence on the context of an XBE*, and *how to determine the alignment between the contexts of two XBEs* have been provided in Chapter 2 while understanding the origin of XBEs. These insights have been used throughout this dissertation to explain the ideas behind the solutions proposed to handle XBEs.

- A detailed description on how the state of the art in IFSs can be used for handling XBEs has been provided in Chapters 3, 4 and 5.

  - In Chapter 3, it has been explained how a collection of XBEs can be characterized as an IFS while answering Research Question Q1. Together with that explanation, a novel approach to compare any two of such IFSs has been described while answering Research Questions Q2, Q3 and Q4. In that approach, the XBEs of some relevant objects are used to build a *connotation-differential print*, which constitutes a representation of a possible difference in the understandings that two persons might have about the topic under evaluation.
  - In Chapter 4, an innovative experimental test procedure has been proposed while answering Research Questions Q2, Q3, Q4 and Q5. In this procedure, similarity measures designed to compare IFSs are tested to determine how suitable these measures are for comparing XBEs characterized as IFSs.
  - In Chapter 5, it has been proposed a novel open-source software package, named *IFSMetrics*, which implements the test procedure described in Chapter 4. With this package, one or more (configurations of) similarity measures can be tested with a big number of

IFSs characterizing XBEs that result from different learning scenarios. In this regard, *IFSMetrics* constitutes an integrated package that helps to answer Research Questions $Q2$, $Q3$, $Q4$ and $Q5$.

- A detailed description on how the novel *augmented appraisal degrees* can be used for handling XBEs has been given in Chapters 6, 7, 8 and 9.
  - In Chapter 6, the novel definition of *augmented appraisal degrees* (AADs), as well as the novel definition of *augmented (Atanassov) intuitionistic fuzzy sets* (AAIFSs) have been proposed while answering Research Question $Q1$. The novel *'as seen from'* operator, the novel *connotation alikeness factor* (CAF) and several novel similarity measures that compare AAIFSs have also been proposed in Chapter 6 while answering Research Questions $Q2$, $Q3$ and $Q4$.
  - In Chapter 7, the novel *k-well-(un)fitted-specimens method* has been proposed while answering the question *how to handle context in XBEs in which only appraisal levels are available*, as well as Research Questions $Q4$ and $Q5$.
  - In Chapter 8, the novel *post-digest method* has been proposed while answering the question *how to handle XBEs when there is no explicit evaluation request,* as well as Research Question $Q4$.
  - In Chapter 9, the novel *flexible attribute-set consensus reaching process* has been proposed, while providing a practical answer to Research Question $Q5$, as an application of AADs to reach consensus on collective XBEs.

10.3.2 Contributions per Research Question

The main contributions of this dissertation while answering the stated research questions are summarized as follows:

Q1. *How to characterize subjective, imprecise and potentially marked-by-hesitation XBEs in such a way that they are suitable for computation?*

Although the description on how to characterize XBEs as elements of an IFS (see Chapter 3) is an important contribution of this dissertation, the main contributions in this regard are the definitions of the AADs and the AAIFSs (see Chapter 6). The AADs and AAIFSs are proven to be (mathematical) representations that are suitable for characterizing subjective, imprecise and potentially marked-by-hesitation XBEs and, also, that lend themselves to computation.

Q2. *How to perform a reliable comparison between XBEs given by a heterogeneous group of respondents?*

With respect to XBEs characterized as elements of an IFS, the definition of the *connotation-differential prints* (CDPs) and the similarity measures that have been *augmented* with such CDPs (see Chapter 3), as well as the open-source software package *IFSMetrics* (see Chapter 5) and the innovative experimental test procedure which has been implemented in this software package (see Chapter 4), are important contributions of this dissertation. While a CDP can help to perform a straightforward but adequate comparison between IFSs characterizing XBEs given by people with different background, *IFSMetrics* can be used for testing such comparisons with IFSs that result from simulated XBEs given respondents who learned a concept under different scenarios.

Regarding XBEs characterized as elements of an AAIFS, the main contributions in this regard are the concepts, operators and methods included into the novel *augmented framework* proposed for comparing AAIFS that represent XBEs given by a heterogeneous group of respondents (see Chapter 6). Among them, a *connotation alikeness factor* and *comparison operators* like '=' or (its fuzzy variant) *'approximately equal'*, as well as similarity measures designed to compare AAIFSs are proven to be important contributions. As was mentioned in Section 10.1.2, while the CAF concept and the comparison operators can be used for comparing two AADs that respectively characterize two XBEs, the similarity measures can be used for comparing two AAIFSs that respectively characterize two *collections* of XBEs.

Q3. *How to measure the perceived quality of XBEs according to a particular understanding?*

The insights about *what aspects might have an influence on the context of an XBE*, and *how to determine the alignment between the contexts of two XBEs* are important contributions in this regard. By applying these insights, the usability and, thus, the perceived quality of XBEs have been linked to their relevance. Since the relevance of the XBEs given by a respondent depends on how aligned his/her understanding of the concept under evaluation is in relation to the understanding possessed by a requester, the definitions of the CDPs and the CAFs (and, so, the other contributions related to the previous research question) become *indirect* contributions that result while answering this question.

Q4. *How to identify XBEs given by (anonymous) respondents with whom a requester shares a similar understanding of the topic under analysis?*

Both AAIFS framework and the regular IFS framework with *augmented* similarity measures, as well as the open-source software package *IFSMetrics* and the experimental test procedure implemented on it are significant contributions in this regard. As was mentioned in Section 10.1.3, both frameworks are viable options to differentiate XBEs that are relevant according to a particular understanding.

Q5. *How to detect and manage automatically any difference in understanding of a concept behind an evaluation request, in which the answers could be given by respondents with different background?*

Even though the *post digest method* and the *flexible attribute-set consensus reaching process* are important practical contributions, the AAIFS framework along with the *k-well-(un)fitted-specimens method* are the most significant contributions in this respect. Both the AAIFS framework and the *k-well-(un)fitted-specimens method* are proven to be suitable for handling any difference in understanding of a concept behind an evaluation request, in which the answers could be given by respondents with different experience or knowledge.

10.3.3 Final Conclusion

The three challenges identified in Section 1.1, as well as the five research questions stated in Section 1.5 have been handled in this dissertation. Hence, in our opinion the purpose of this PhD study phrased in Section 1.2 has been achieved.

Index