Argumentation Dialogues in Web-based GDSS:
An approach using Machine Learning Techniques

Luís Manuel da Silva Conceição

UMinho | 2023

Universidade do Minho
Escola de Engenharia

Luís Manuel da Silva Conceição

Argumentation Dialogues in Web-based: GDSS: An approch using Machine Learning Techniques

January 2023

Universidade do Minho
School of Engineering

Luís Manuel da Silva Conceição

Argumentation Dialogues in Web-based GDSS: An approach using Machine Learning Techniques

Doctorate Thesis

Doctorate in Informatics

Work developed under the supervision of:
Professor Doutor Paulo Jorge Freitas de Oliveira Novais
Professora Doutora Maria Goreti Carvalho Marreiros

January, 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

______________________, ______________________________
(Place) (Date)

____
(Luís Manuel da Silva Conceição)

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v6.10.17) [52].

Aos meus pais e irmão.
À memória da avó Graça e do avô João.

Acknowledgements

I want to express my sincere gratitude to my supervisor, Professor Paulo Novais, for his invaluable guidance, support, and encouragement throughout my Ph.D. studies. His expertise, insights, and patience have been instrumental in helping me to complete this research.

To my co-supervisor, Professor Goreti Marreiros, I would be thankful for the rest of my life for all the support, guidance, encouragement, and above all, for the true friendship.

To João, a friend I made since I joined GECAD, for all the knowledge transmitted, teachings, and sermons, and, above all, for the friendship and all the support provided - Thank you!

To my friends and work colleagues, also Ph.D. candidates Jorge and Diogo - THANK YOU! We started this journey together, in different institutions, and we will also finish it ”together”. I wish you all the greatest success, and I will be there as soon as your day arrives!

To my former professors, now co-workers, and above all, friends: Constantino and Ana, thank you for all the support, encouragement, and strength you gave me to complete this stage.

To my three ancient / age-old / obsolete, or simply *”velhotes”* (there are so many synonyms!!), friends - now that it's delivered, let's book lunch at my house! There is also the promise that there will never be a lack of wine at the end of the afternoons! In fact, the best prize of these years is the people I take with me for life! - Thank you for everything, especially for being there!

I am especially grateful to all the students who did their research internships with me, particularly my master students. You helped me become a better supervisor, and without your contributions and motivation, this thesis would not be complete.

My institutional thanks go to Algoritmi and GECAD because, in addition to being a second home throughout all this time (in the case of GECAD), both research units provided me with a good working environment and all the necessary conditions to carry out this work. I also thank to *Fundação para a Ciência e a Tecnologia*, for the Ph.D. grant funding with the reference: SFRH/BD/137150/2018.

Last but not least, I want to thank my family and close friends, special kiss and hug to João, Cláudio, Nádia, and Jorge. To my parents, for everything they have given me throughout my life, for the love and affection they give me. I want to apologize for the times when I complained about everything and nothing in the most complicated moments! To my brother, thank you for always giving me reasons to complain to him! After all, that's what brothers are for!

"Those who can imagine anything, can create the impossible."

— Alan Turing

(Mathematician, Computer Scientist)

Resumo

A tomada de decisão está presente no dia a dia de qualquer pessoa, mesmo que muitas vezes ela não tenha consciência disso. As decisões podem estar relacionadas com problemas quotidianos, ou podem estar relacionadas com questões mais complexas, como é o caso das questões organizacionais. Normalmente, no contexto organizacional, as decisões são tomadas em grupo.

Os Sistemas de Apoio à Decisão em Grupo têm sido estudados ao longo das últimas décadas com o objetivo de melhorar o apoio prestado aos decisores nas mais diversas situações e/ou problemas a resolver. Existem duas abordagens principais à implementação de Sistemas de Apoio à Decisão em Grupo: a abordagem clássica, baseada na agregação matemática das preferências dos diferentes elementos do grupo e as abordagens baseadas na negociação automática (e.g. Teoria dos Jogos, Argumentação, entre outras).

Os atuais Sistemas de Apoio à Decisão em Grupo baseados em argumentação podem gerar uma enorme quantidade de dados. O objetivo deste trabalho de investigação é estudar e desenvolver modelos utilizando técnicas de aprendizagem automática para extrair conhecimento dos diálogos argumentativos realizados pelos decisores, mais concretamente, pretende-se criar modelos para analisar, classificar e processar esses dados, potencializando a geração de novo conhecimento que será utilizado tanto por agentes inteligentes, como por decisiores reais. Promovendo desta forma a obtenção de consenso entre os membros do grupo. Com base no estudo da literatura e nos desafios em aberto neste domínio, formulou-se a seguinte hipótese de investigação - É possível usar técnicas de aprendizagem automática para apoiar diálogos argumentativos em Sistemas de Apoio à Decisão em Grupo baseados na web.

No âmbito dos trabalhos desenvolvidos, foram aplicados algoritmos de classificação supervisionados a um conjunto de dados contendo argumentos extraídos de debates online, criando um classificador de frases argumentativas que pode classificar automaticamente (A favor/Contra) frases argumentativas trocadas no contexto da tomada de decisão. Foi desenvolvido um modelo de clustering dinâmico para organizar as conversas com base nos argumentos utilizados. Além disso, foi proposto um Sistema de Apoio à Decisão em Grupo baseado na web que possibilita apoiar grupos de decisores independentemente de sua localização geográfica. O sistema permite a criação de problemas multicritério e a configuração das preferências, intenções e interesses de cada decisor. Este sistema de apoio à decisão baseado na web inclui os *dashboards* de relatórios inteligentes que são gerados através dos resultados dos trabalhos alcançados pelos modelos anteriores já referidos. A concretização de cada um dos objetivos permitiu validar as questões de investigação identificadas e assim responder positivamente à hipótese definida.

Palavras-chave: Ciência de Computadores, Inteligência Artificial, Diálogos Argumentativos, Sistemas de Apoio à Decisão em Grupo

Abstract

Decision-making is present in anyone's daily life, even if they are often unaware of it. Decisions can be related to everyday problems, or they can be related to more complex issues, such as organizational issues. Normally, in the organizational context, decisions are made in groups.

Group Decision Support Systems have been studied over the past decades with the aim of improving the support provided to decision-makers in the most diverse situations and/or problems to be solved. There are two main approaches to implementing Group Decision Support Systems: the classical approach, based on the mathematical aggregation of the preferences of the different elements of the group, and the approaches based on automatic negotiation (e.g. Game Theory, Argumentation, among others).

Current argumentation-based Group Decision Support Systems can generate an enormous amount of data. The objective of this research work is to study and develop models using automatic learning techniques to extract knowledge from argumentative dialogues carried out by decision-makers, more specifically, it is intended to create models to analyze, classify and process these data, enhancing the generation of new knowledge that will be used both by intelligent agents and by real decision-makers. Promoting in this way the achievement of consensus among the members of the group. Based on the literature study and the open challenges in this domain, the following research hypothesis was formulated - It is possible to use machine learning techniques to support argumentative dialogues in web-based Group Decision Support Systems.

As part of the work developed, supervised classification algorithms were applied to a data set containing arguments extracted from online debates, creating an argumentative sentence classifier that can automatically classify (For/Against) argumentative sentences exchanged in the context of decision-making. A dynamic clustering model was developed to organize conversations based on the arguments used. In addition, a web-based Group Decision Support System was proposed that makes it possible to support groups of decision-makers regardless of their geographic location. The system allows the creation of multicriteria problems and the configuration of preferences, intentions, and interests of each decision-maker. This web-based decision support system includes dashboards of intelligent reports that are generated through the results of the work achieved by the previous models already mentioned. The achievement of each objective allowed validation of the identified research questions and thus responded positively to the defined hypothesis.

Keywords: Computer Science, Artificial Intelligence, Argumentation Dialogues, Group Decision Support Systems

Contents

List of Figures

List of Tables

Acronyms

CRP	Consensus Reaching Process
FCT	Fundação para a Ciência e a Tecnologia
FMUP	Faculdade de Medicina da Universidade do Porto
GDM	Group Decision-Making
GDSS	Group Decision Support System
ISEP	Instituto Superior de Engenharia do Porto
LSGDM	Large-Scale Group Decision-Making
LSTM	Long Short-Term Memory
MAS	Multi-Agent System
ML	Machine Learning
MPMCDM	Multi-Person Multi-Criteria Decision-Making
NLP	Natural Language Processing
POI	Point of Interest
PRH	Predictive and Relevance-Based Heuristic agent
PROMETHEE	Preference Ranking Organization Method for Enrichment Evaluation
PRR	Plano de Recuperação e Resiliência
SVM	Support Vector Machine
SWOT	Strength, Weakness, Opportunity and Threat
TOPSIS	Technique for Order Preference by the Similarity to Ideal Solution

UMAP	Uniform Manifold Approximation and Projection
WBAF	Weighted Bipolar Argumentation Framework

Chapter 1

Introduction

Decision-making is present in every person's day-to-day life, even if they often do not perceive it [76]. Decisions may be related to everyday problems, such as the choice of the next day's wake-up time, or they may be related to more complex issues, including organizational issues, such as the acquisition of a new manufacturing plant for a multinational company. The first case refers to a decision that concerns a single individual; in the second, this process is usually executed by a group of people, adopting the denomination of the group decision-making process [17]. Decision-making is also present, and with a greater impact, in society in general, from the state, the government, public institutions, and private organizations.

With regard to the size of the groups, and given the difficulty in accurately characterizing this variable, DeSanctis and Gallupe [27] opted for the designations "small"and "large". The literature of the area usually describes groups until 5 people as considered "small", groups from 6 to 12 people as "medium", and groups with more than 12 people as large. The proximity variable between group members represents the degree of contiguity, in terms of space and time, that exists between them. More recently, Garcia-Zamora et al. [37] consider that the concept of Large Groups can encompass groups of thousands of decision-makers.

Group Decision Support System (GDSS) have been studied over the last decades with the goal of improving the support provided to decision-makers in many different situations and/or problems to solve. There are two main approaches to the implementation of Group Decision Support Systems: the classical approach, based on the mathematical aggregation of the preferences of the different elements of the group, and the approach based on automatic negotiation. Approaches based on the aggregation of preferences, usually called multi-criteria analysis, combine decision-maker's preferences to select an alternative or order a set of alternatives [10, 41, 75]. Approaches based on automatic negotiation can have their genesis in game theory, heuristic approaches, or argumentation [39, 70, 78]. In the case of approaches based on game theory or heuristic procedures, decision-makers exchange proposals and counter-proposals until a solution is constructed that is accepted by both parties. In the case of argumentation-based approaches, it is possible to exchange explanations or justifications about, for example, a certain alternative or criterion, which can facilitate reaching agreements since it enhances the expansion of knowledge and cognitive abilities of the decision-maker.

Current argumentation-based GDSS can generate a huge amount of data [8, 29, 49]. The aim of this research work is to study and develop models using machine learning techniques to extract knowledge from argumentative-based dialogue performed by decision-makers. Particularly, it is intended to model argumentative processes in GDSS, using multi-agent systems, considering the decision-makers' objectives and understanding the decision process. This work intends to create models to analyze, classify and process this data to potentiate the generation of new knowledge that will be used by both agents and decision-makers.

This chapter presents to the reader an overview of the work that has been developed. It begins with a contextualization of the topics under which this work was developed, namely Section 1.1 introduces the concepts related to group decision-making. Following, in Section 1.2, some concepts related to GDSS and some of the recently published works are presented. Section1.3 presents published works that use Machine Learning techniques applied to the topic of group decision-making. The hypothesis raised for this research work is described in Section1.4, followed by the Objectives in Section1.5. Finally, the applied Research Methodology is explained in Section1.6, and in the last Section1.7 a description of the full structure of this Doctoral Thesis is presented.

1.1 Group Decision-Making

Group decision-making is a process through which a group of people, called participants (decision-makers), interact with each other, analyzing a set of variables and considering and evaluating the available alternatives with the objective of selecting one or more solutions for a given problem [57].

Multi-criteria decision problems are characterized by two or more alternatives, which in turn are composed of two or more criteria that may conflict with each other. This type of problem tends to be quite complex, and its resolution requires a structured and logical decision-making process. Although complex, these problems can be found in everyday life or in strategic decisions of organizations. In fact, according to Saaty [76], choosing one among several alternatives is a natural process in human beings. There are several factors that hinder the process of establishing preferences, and that contribute to increasing its complexity [40], namely the fact that the criteria are often conflicting, the number of alternatives under evaluation is very high and that most of the time the decision-maker does not have complete information, making it necessary to reason with uncertainty.

Finally, it is important to remember that in the context of group decision-making, it is intended that the multi-criteria problem be analyzed and solved by a group of individuals, which with some frequency has a multidisciplinary background, transforming the obtainment of a final solution in an even more complex process. The work developed in this thesis will focus on supporting groups of decision-makers on the path to finding solutions for multi-criteria problems.

There are a number of reasons why decisions are taken in a group, whether they are: improving decision quality, sharing responsibilities, gaining support among stakeholders, training less experienced members, and obviously, due to the organigrams of today's companies that are obliged to do so [44, 45]. However, there are also disadvantages associated with group decision-making, such as the pressure of compliance, fear of evaluation, the dominance of the discussion by one of the elements, and carrying out incomplete analyses. These advantages and disadvantages do not necessarily have to coexist, i.e., to manifest themselves simultaneously. Its existence often depends on the size of the group, the nature of the task, the type of meeting, and the participants. The number of participants involved in the process is variable and may all be in the same location at the same time or geographically dispersed at different times (as represented in Figure 1) [53, 63].

During the last decade, a new concept emerged: The Large Scale Group Decision-Making. It consists of a process of making decisions by a group of people, typically in organizations or societies, where the group is large enough that traditional face-to-face communication and decision-making processes are not practical. This type of decision-making can be used to solve problems, make plans, or set policies in a variety of settings, such as business, government, and community organizations[29].

The quality of the decisions made is one of the most important factors determining the organization's success [53]. That is why supporting a decision-making process is a complex task, mainly when the use of automation mechanisms is intended. Automatic negotiation, or more specifically argumentation, has been widely used in GDSS to automate negotiation processes [12].

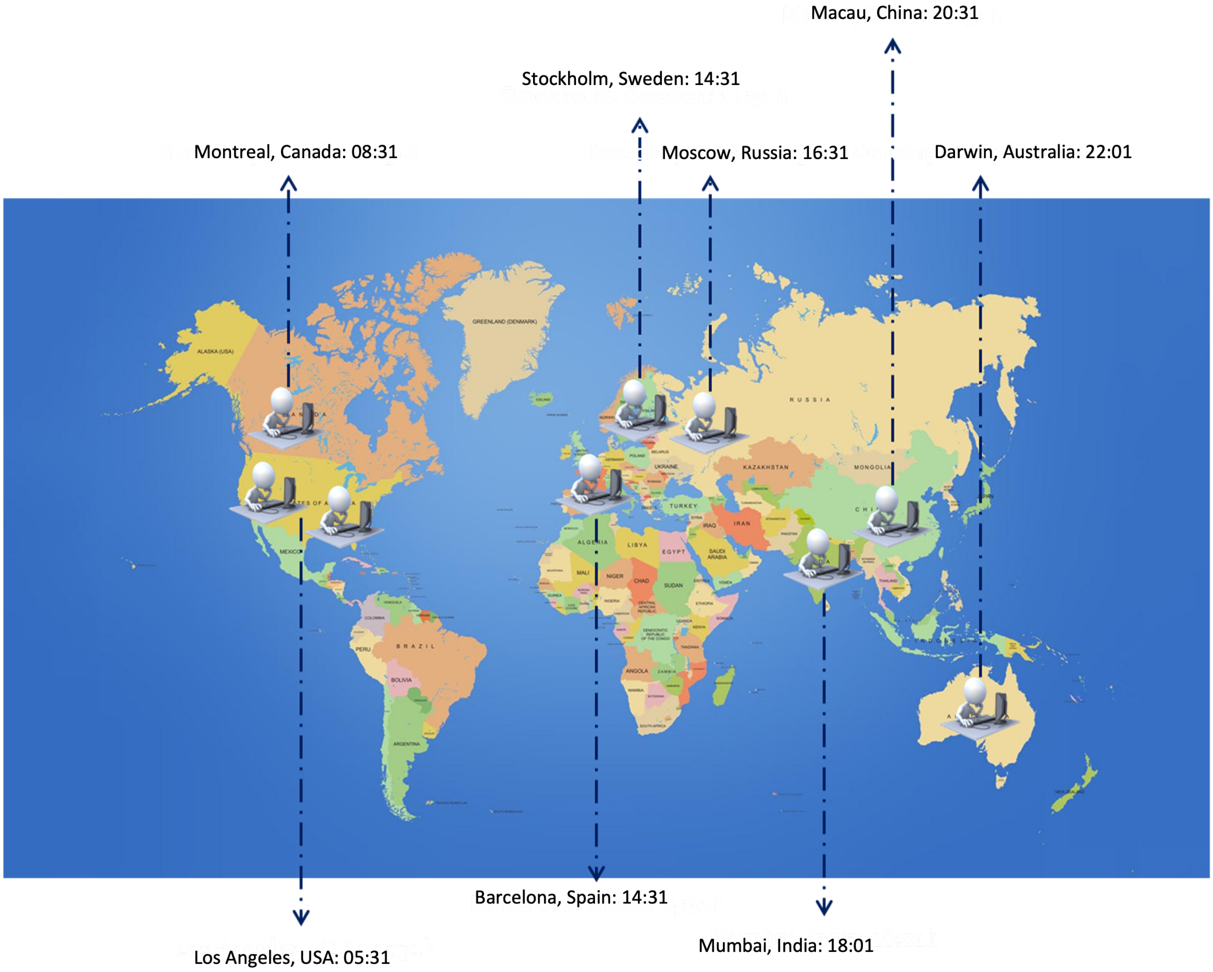

Figure 1: Dispersed Decision-Makers in different time zones (adapted from [24])

In the literature, it is possible to find several approaches to large-scale group decision-making including voting systems, multi-criteria decision analysis, negotiation, and consensus building. Each approach has its own strengths and weaknesses, and the most appropriate method to incorporate in a Large Scale GDSS will depend on the specific context and goals of the decision-making process.

1.2 Group Decision Support Systems

The "Group Decision Support System"concept emerged in the 1980s when Huber [44] defined it as a combined structure of software, hardware, languages, and procedures that support the work of a group whose aim is the purpose of carrying out a decision-making process. In this section (1.2) some concepts of the foundation of GDSS will be presented in 1.2.1. Then, in 1.2.2 some features of a few of the most relevant GDSS will be described. The last section of this chapter is about Large-Scale Group Decision-Making (LSGDM).

1.2.1 Background

Huber [44] defined three essential activities as an integral part of a GDSS: information gathering, information sharing and information use. The collection of information concerns the selection of data from a database or simply the collection of group members' opinions. The sharing of information is related to data availability to decision-makers in the decision group. The use of information refers to the creation of applications using procedures and group problem-solving techniques in order to reach a decision.

DeSanctis and Gallupe [27], also identified, in 1984, the concept of GDSS and referred that "an exciting new concept is emerging in the area of decision support". According to the authors, a GDSS is an interactive computer system that helps a group of decision-makers solves unstructured problems. They proposed five key features for a GDSS:

- GDSS are systems specifically developed and not a simple configuration of some existing components;
- its purpose is to assist groups of decision-makers in their work, improving the decision-making process as well as the results of it;
- should be easy to learn and easy to use;
- can be specific, developed for a particular type of problem, or general if developed to assist in making different types of decisions in an organization;
- should include internal mechanisms that minimize the development of negative behaviors in the group of decision-makers, such as the destructive concept or lack of communication.

These two authors carried out remarkable work in the study of GDSS, proposing a multi-dimensional taxonomy for their characterization based on the following dimensions: the type of task, the proximity between the members of the group, and the size of the group (as illustrated by the Figure 2).

Decision room (small group and face-to-face): the decision-makers are all in the same physical space, meeting for a certain period of time to discuss a problem or a set of problems. The meeting is held in a room with a screen to project the information, and each decision-maker has a computer. Communication between members is transmitted verbally or via electronic messages. Local decision network (small and dispersed group): in this scenario, there are several possible configurations; in the first, the group members are located at workstations in offices and are connected through a local network; in the second the members who are at home or traveling can connect through long-distance networks; and in the third, two or more decision rooms interconnected by teleconference. This type of meeting can last longer than a day, and it is not necessary for all group members to be online simultaneously. Legislative session (large group and face-to-face): the big difference to the decision room is that, in this case, only the facilitator can send information to the shared screen. Individual computers can be shared by groups of 2 or 3 people. Communication occurs in a hierarchical way, where each decision-maker can send

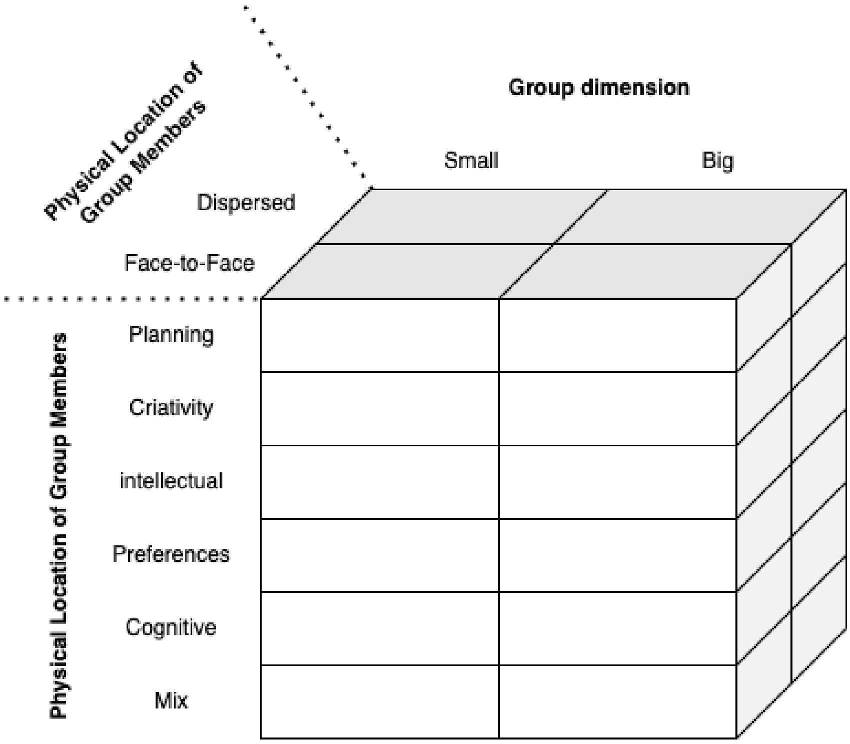

Figure 2: Multi-Dimensional taxonomy for study of GDSS (adapted from [28])

messages only to decision-makers who are at the same level as theirs or to the member responsible for that group. Electronic conference (large and dispersed group): when a large number of people are not physically close and need to come together to make decisions, a different GDSS is needed. Long-distance telecommunications networks are required. In addition, there must be a hierarchy of decision-makers, as in legislative sessions [28].

Regarding the size of the groups, and given the difficulty in accurately characterizing this variable, the authors opted for the designations “small” (groups of 3 to 5 people) and “large” (groups of more than 12). The proximity variable between group members represents the degree of contiguity, in terms of space and time, that exists between them.

Desanctis and Gallupe (1987) also defined a taxonomy that considers four configurable environments according to the size of the group and the dispersion of its elements. These factors are not mutually exclusive but represent bipolar extremes that influence meeting dynamics and decision-support feature selection. The four environments are characterized in Table 1.

Table 1: GDSS taxonomy - Physical Location vs Group Dimension (adapted from [28])

		Group Dimension	
		Small	**Big**
Physical Location of Group Members	Face-to-Face	Decision room	Legislative session
	Dispersed	Local decision network	Electronic conference

1.2.2 Web-based Group Decision Support Systems

GDSS have evolved over the years [8, 19, 24, 55–57], and it is possible to find several GDSS that are web-based to support decision-makers in many areas of society, such as healthcare, economy, gastronomy, logistics, and industry. An example, Miranda et al. [59] proposed a simulated medical practice scenario to deal with staging cancer. In their proposal, the decisions were made in groups and allowed collaborative work. These authors also implemented a Multi-Agent System (MAS) to represent and exchange information related to real participants [59].

In work proposed by Tavana et al. [82], a GDSS was developed to evaluate and manage oil and natural gas transportation using the alternative pipeline routes from the Caspian Sea to other regions. They used the Delphi method to represent decision-maker's beliefs using the Strength, Weakness, Opportunity and Threat (SWOT) analysis. These beliefs were integrated using the Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) in order to find a better solution for pipeline routes [82].

In the work proposed by Morente-Molinera et al. [60], they developed a decision support system composed of a web and a mobile application to support the selection of wines. This system allowed decision-makers to participate in the decision-making process even if they were geographically dispersed. They considered using different techniques, such as a fuzzy wine ontology, group decision support algorithms, and a fuzzyDL reasoner [60].

Yazdani et al. [94] proposed a group decision support system for selecting logistic providers. Their model combines a Quality Function Deployment and the multi-criteria decision analysis algorithm Technique for Order Preference by the Similarity to Ideal Solution (TOPSIS) to optimize a French logistic agricultural distribution center. They proposed a model that approaches the decision problem considering the technical and the customer perspectives. The system acts as an interface between decision-makers and customer values to select third-party logistic providers. The system uses fuzzy linguistic variables to support agricultural parties in uncertain situations [94].

Regardless of the potential offered by web-based GDSS, the success and acceptance of these systems have not been positive by the decision-makers so far. Some known reasons are related to the resistance to change from employees and difficulties in the configuration of the system, either in the creation and configuration of the decision problem or in the configuration of the preferences for each decision-maker. Another reason that makes web-based GDSS hard to accept is the fear of losing dialogue, and ideas discussion that can be achieved in face-to-face meetings [13, 42, 43].

Ghavami, Taleai, and Arentze [38] proposed an intelligent web-based spatial group decision support system to investigate the role of opponents modeling in urban land use planning by using a multi-agent system approach. The system was designed to support the analysis of various land use scenarios and to facilitate communication among stakeholders, including planners, policymakers, and community members. They implemented a multi-agent system that helps decision-makers (stakeholders) to reach a consensus regarding urban land planning [38].

1.2.3 Large-Scale Group Decision-Making

The LSGDM area started to be studied in the last decade, and the number of publications is increasing year after year [29, 80], being considered a hot topic in the decision science domain. According to Ding et al. [29], most current approaches consist of extending classic approaches that have been used for small groups. Due to their nature, LSGDM events are more complex than "traditional" group decision-making processes since it is necessary to deal with new challenges in addition to the ones already considered before. The Consensus Reaching Process (CRP) is an essential process in LSGDM events since in most cases it is impossible to reach a consensual decision after the participants' initial preferences configuration [29]. It consists of a dynamic feedback process that aims to reduce existing conflicts as the group seeks to converge toward a solution [49]. Most existing works consider that a decision becomes consensual when a certain level of Consensus Degree is reached [10, 31]. However, reaching a consensus can be an extremely complex process, as the participants involved can be highly diverse (differing in their intentions, preferences, motivations, level of expertise, etc. [8]). There are 4 strategies that are typically used in CRP to support and enhance consensus:

- Clustering phase consists of the creation of clusters of participants according to common characteristics. Tang and Liao [80] divided the main methods that have been applied in the context of LSGDM for creating clusters in 3 groups: hierarchical clustering methods (e.g., Agglomerative Hierarchical Clustering algorithm [88, 89]), partitioning clustering methods (e.g., K-means algorithm [81] and Fuzzy Equivalence Relation algorithm [90]) and other clustering methods (e.g., Self-Organizing Maps [93] and Partial binary tree DEA-DA cyclic model[50]). For the creation of clusters, researchers have considered aspects such as: the distance to the opinion collection, the decision-makers' preferences, and relationships [29];

- Detection of non-cooperative participants: consists of identifying participants or clusters that are hindering a consensus. Most identification approaches are based on distance measures (e.g., the distances between clusters centers and the collective opinion [51], or the consensus levels within clusters [91, 92]), but there are others based on the degree of cooperativeness [67, 68] and the degree of conflict (e.g., opinion conflict and behavior conflict [30]);

- Feedback phase: consists of suggesting changes participants should make in their assessments in order to promote consensus [49]. The main approach is to recommend participants who were found to be non-cooperative change their assessments to bring them closer to the aggregated collective opinion [62];

- Weight updating phase: consists of updating the weights assigned to participants or clusters identified as non-cooperative. The main strategy is to implement mechanisms that penalize the weight that these decision-makers or clusters have in the decision [51, 64].

In the context of this research work, these techniques can be applied to support the identification of different types of non-cooperative behaviors; the identification, in the conversations, of arguments in favor and against the topic under discussion; and the conception of intelligent strategies to enhance consensus from the decision quality perspective.

1.3 Machine Learning Approaches in Argumentation-based Dialogues

The volume of information produced by dialogues during a decision-making process makes it difficult to extract knowledge that is relevant in the context. In the literature, there are approaches that combine argumentation with Machine Learning (ML) techniques.

1.3.1 Argumentation Dialogues

Argumentation, as an area of study, is of great complexity and incorporates the study of transversal areas such as philosophy, communication, linguistics, and psychology [84], being widely used in Computer Science and in Artificial Intelligence. The first works developed in argumentation in the field of Artificial Intelligence date back to the beginning of the 1980s [5, 36], being specifically related to the study of natural language processing. The group decision-making process involves multiple decision-makers, possibly with different past experiences, preferences, and personal characteristics (e.g. personality, affective state), and consequently with different perspectives of the problem under consideration. This process enhances the occurrence of the most different forms of conflict that must be overcome through interaction and dialogue to reach a final consensual solution.

Walton and Krabbe [87] proposed a taxonomy in which they classify dialogues based on their main objective, the objectives of each of the participants, and the initial knowledge of each participant. Argumentation plays a key role in different types of dialogues, allowing the use of justifications and explanations to influence the preferences of other decision-makers and, consequently, the outcome of the decision-making process [2, 54, 70]. In the 90s, several models of argumentation were proposed, with great impact in the area of computer science and, in particular in the area of Multi-Agent Systems [32, 46, 77]. In recent decades, a significant number of new argumentation models have emerged, in some cases, extensions of existing models: abstract argumentation [66], logic-based argumentation [1], value-based argumentation [25], argumentation based on assumptions [35], among others [3].

The argumentation models proposed over the years, and based on the classification of Walton and Krabbe's dialogues, focus essentially on negotiation and persuasion. For example, Kraus, Sycara, and Evenchik [46] describes a logical model for reaching agreements through argumentation in which the arguments exchanged between the parties are based on the psychology of persuasion. However, analyzing the description of the objectives of the dialogues, the deliberation, whose objective is the choice of a final solution, proves to be very suitable for group decision processes. Despite that, there is still a lack of models that support this type of dialogue. One of the great challenges is the cooperation between decision-makers and intelligent agents. Currently, the existing models focus on dialogues between agents, not considering the important role that human decision-makers have in the process [4, 69, 83]. An initial approach to this problem was proposed in 2018 by Carneiro et al. [12]. In their work, an argumentation model based on dialogues was proposed that allows decision-makers and intelligent agents to share the knowledge that is generated throughout the decision process. Decision makers can evaluate received messages, which can be used by agents to understand the attack/reinforcement relationship between messages [12]. The cooperation between intelligent agents and decision-makers is a key success factor for overcoming the challenges of not having face-to-face meetings.

1.3.2 Related works

It is possible to find different approaches in this area; for instance, in the work of Rosenfeld and Kraus [72] they used ML techniques in a multi-agent system that supports humans in argumentative discussions by proposing possible arguments to use. To do this, they analyzed the argumentative behavior of 1000 human participants, and they proved that it is possible to predict human argumentative behavior using ML techniques. In their results, they demonstrated that ML techniques could achieve up to 76% accuracy when predicting people's top three argument choices given a partial discussion. Their MAS implementation has 9 argument provision agents, which they empirically evaluated using hundreds of human study participants. The Predictive and Relevance-Based Heuristic agent (PRH), uses ML prediction with a heuristic that estimates the relevance of possible arguments to the current state of the discussion, resulting in significantly higher satisfaction levels among study participants compared with the other evaluated agents. These other agents propose arguments based on Argumentation Theory, propose predicted arguments without the heuristics or with only the heuristics, or use Transfer Learning methods. Their findings also show that people use the PRH agents proposed arguments significantly more often than those proposed by the other agents [72].

Another interesting work is the one published by Carstens and Toni [15] that deals with the identification and extraction of argumentative relations. To do that, they built a corpus and classified pairs of sentences according to whether they stand in an argumentative relation to other sentences, considering any sentence as argumentative that supports or attacks another sentence. They used two types of features as input for their ML model: Relational features and Sentential features. Relational features they have considered are the ones that represent how the two sentences that make up the pair relate to each other like WordNet-based similarity [58], Edit Distance measures [61], and Textual Entailment measures [26]. In the Sentential features category, they include a set of features that characterize the individual sentences, such as various word lists (keeping count of discourse markers), sentiment scores (using SentiWordNet [34]) or the Stanford Sentiment library [79], among others. Their experiments on the preliminary corpus, representing sentence pairs using all features described, showed good results, with a classification accuracy of up to 77.5% when training Random Forests on the corpus [15].

In Cocarascu and Toni [18], they propose a deep learning architecture to identify argumentative relations of attack and support from one sentence to another using Long Short-Term Memory (LSTM) networks. The proposed architecture uses two LSTM networks (unidirectional and bidirectional) and word embeddings. Their unidirectional LSTM model with trained embeddings and a concatenation layer achieved 89.53% accuracy and 89.07% F1. They affirm that their results indicate that LSTMs may be better suited for Relation-based Argument Mining at least for non-micro texts [6] than standard classifiers as used in e.g. [16], as LSTMs are better at capturing long-term dependencies between words and they operate over sequences, as found in text. They conclude, considering that their work allowed to considerably improve upon existing techniques that use syntactic features and supervised classifiers for the same form of (relation-based) argument mining [18].

Authors Rosenfeld and Kraus [74] presented a novel methodology for automated agents for human persuasion using argumentative dialogues [73, 74]. This methodology is based on an argumentation framework called Weighted Bipolar Argumentation Framework (WBAF) and combines theoretical argumentation modeling, ML, and Markovian optimization techniques, resulting in an innovative agent named SPA. They performed field experiments and concluded that their methodology enabled the automated agent, SPA, to persuade people in 2 distinct environments. In both an attitude change environment and a behavior change environment, SPA was able to perform on a human-like level and significantly better than a baseline agent. They concluded that in their experiences, an agent could persuade people no worse than people can persuade each other [73, 74].

In Zuheros et al. [95], the authors propose a Multi-Person Multi-Criteria Decision-Making (MPMCDM) methodology that uses sentiment analysis and deep learning to aid in decision-making. The case study presented in the paper uses this methodology to help individuals choose a restaurant using TripAdvisor reviews. To implement the MPMCDM methodology, the authors first used Natural Language Processing (NLP) techniques to extract the sentiment from the TripAdvisor reviews. They then used deep learning algorithms to analyze the sentiment and make recommendations based on the preferences of the individuals involved in the decision. Overall, the authors propose that this MPMCDM methodology can be used as a decision aid to help individuals make smarter decisions by taking into account the sentiment of others. The case study of restaurant choice using TripAdvisor reviews demonstrates the potential practical applications of this methodology [95].

Lawrence and Reed [47] elaborated a survey describing several approaches to argument mining that use machine learning algorithms. Some of these approaches involve training a machine learning model on a labeled argumentative text dataset, where human experts have manually identified and annotated arguments. The model can then be used to identify and classify arguments in the new, unseen text.

The most common types of machine learning algorithms that have been used for argument mining include [47]:

- Supervised learning algorithms: These algorithms require a labeled training dataset, in which the input data (e.g., text documents) and the corresponding output labels (e.g., argumentative or non-argumentative) are provided. The algorithm learns a function that maps the input data to the correct output labels. Examples of supervised learning algorithms that have been used for argument mining Support Vector Machine (SVM) and decision trees.

- Unsupervised learning algorithms: These algorithms do not require a labeled training dataset. Instead, they try to identify patterns and relationships in the data through techniques such as clustering and dimensionality reduction. Unsupervised learning algorithms have not been widely used for argument mining, but they have been applied in some cases to identify argumentative structures within the text.

- Semi-supervised learning algorithms: These algorithms use a combination of labeled and unlabeled data to learn the function that maps inputs to outputs. They can be useful in cases where it is difficult or expensive to obtain a large labeled training dataset, as they can use both, labeled and unlabeled data to improve the accuracy of the model.

1.4 Motivation and Research Problem

In the previous sections of this thesis, a general overview of Group Decision-Making and Group Decision Support Systems has been provided along with some limitations and constraints in the adoption of these systems. Supporting a group decision-making process is a complex task, mainly when the automation of mechanisms is intended. As was stated, group decision-making tasks can be very time-consuming for the decision-makers, and the process itself, even if supported by a GDSS, can struggle with the decision process. This impact in the process can even be worse when the group dimension fits the LSGDM paradigm. The complexity of the group decision-making process is directly related to the complexity of the decision problem, but also to the number of involved participants. In the LSGDM processes, where the number of participants can reach thousands, the need to structure and extract information exponentially increases. Either in the scenario of LSGDM or smaller groups, argumentation has been widely used in GDSS to support negotiation processes, and it is a fact that argumentation dialogues generate a huge quantity of data.

This opens the floor to the conception and the development of new methods and models able to analyze, classify and process the information that is exchanged between decision-makers, through the argumentation process, to potentiate the generation of new knowledge that can be used to facilitate the obtention of consensus among the group members. To address this issue, this research work aims to study and develop models using machine learning techniques to extract knowledge from argumentative-based dialogue models performed by decision-makers. This work intends to create models to analyze, classify and process this data to create conditions so that the participants understand the reasons that lead to a certain decision, receiving personalized information throughout the process while encouraging new contributions. Based on the literature study, and the open challenges existent in this domain, the following research hypothesis was formulated:

> ***It is possible to use Machine Learning techniques to support argumentation dialogues in web-based Group Decision Support Systems***

Based on this hypothesis, the following research questions were identified and will guide the execution of the work:

- Is it possible to perform an automatic assessment of the arguments introduced by real participants in a decision-making process?
- Applying Machine Learning techniques allows to dynamic generate clusters of the messages exchanged by decision-makers enabling them to receive/access personalized information throughout the process while encouraging new contributions?
- Is it possible to conceive and develop a web-based GDSS prototype based on a multi-agent system that considers the valuable intelligence and knowledge that can be generated in face-to-face meetings?

1.5 Objectives

The research questions previously formulated enabled the definition of the main goal of this research work: to study and develop models using machine learning techniques to extract knowledge from argumentative-based dialogue models performed by decision-makers. To achieve this goal, the following objectives are proposed:

- Study and develop a model that allows the system itself to assess the arguments introduced by real participants automatically;
- Study and develop models that will be able to dynamically generate clusters of the messages exchanged by decision-makers.
- Design a web-based GDSS prototype based on a multi-agent system that will include the models previously described.

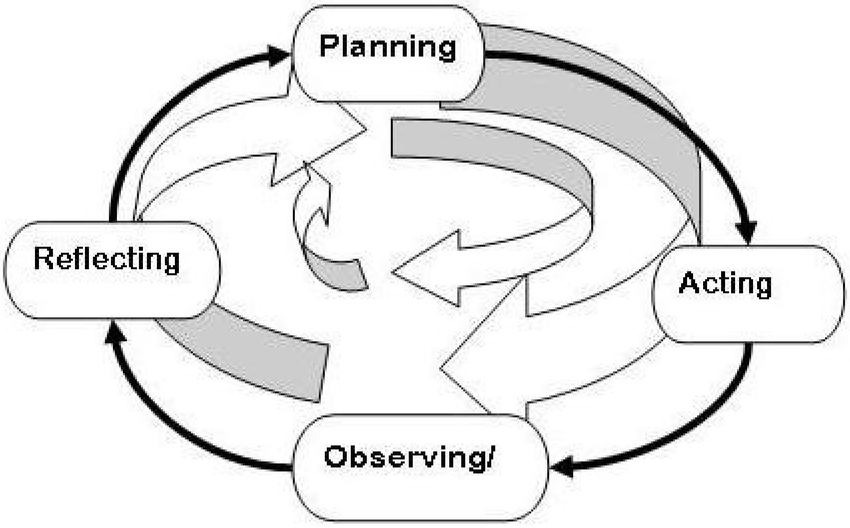

Figure 3: Action Research model [48]

1.6 Research Methodology

In order to achieve the defined objectives, the adopted approach is deductive-inductive, where several conceptual theories, tests, and practical solutions are analyzed. In what concerns the research strategy, it was used Action Research methodology.

Action research involves actively participating in and reflecting on the actions and experiences of a group or community to effect change and improve a specific situation (as illustrated in Figure 3). It is typically conducted in a collaborative and participatory manner, with the goal of achieving a greater understanding of the problem or issue being studied and developing practical solutions to address it. Action research is often used in fields such as education, social work, and community development, but it can be applied in a wide range of settings. It is particularly well-suited for addressing complex and multifaceted problems that require the involvement and participation of various stakeholders [48].

1.7 Structure of the document

This section presents the structure of the thesis, and it is intended to be a guide to help readers find content and explain some aspects of the document's structure. This thesis follows a format called "Compilation Based"or "Scandinavian model", which consists of the compilation of several publications that present the methods and results regarding the initially defined hypotheses. The structure of the three chapters that compose the document is the following:

Chapter 1: Introduction

This chapter 1 starts by describing the context of this work and the motivation, and it is intended to lead the reader to understand the relevance of the work. Section 1.1 contextualizes the reader about the Group Decision-Making concepts and scope that will be referred forward in the document. Next, Section 1.2

provides a brief history of GDSS and their evolution to the current times. Then, Section1.3 presents published works that use Machine Learning techniques applied to the topic of group decision-making. The Hypothesis raised for this research work is described in Section1.4, followed by the Objectives in Section1.5. Following, the applied Research Methodology is explained in Section1.6, and finally, this chapter ends with a description of the structure and contents present in this thesis.

Chapter 2: Publications Composing the Doctoral Thesis

This chapter presents a set of four papers that have been developed during this research work. Before the presentation of the publication itself, a table is presented describing all the information regarding the publication.

- *Section 2.1: A web-based group decision support system for multicriteria problems [22]*
  This paper describes the development of a web-based Group Decision Support System (GDSS) for supporting multi-criteria decision-making processes in dispersed groups. The system allows decision-makers to create multi-criteria problems and to specify their preferences, intentions, and interests. Then this information is combined and processed by virtual agents using a Multi-Agent System (MAS). This approach preserves the valuable intelligence and knowledge that can be generated in face-to-face meetings and has a high level of usability, which may contribute to the easier acceptance and adoption of GDSS. The paper discusses the design and implementation of the system, as well as the workflow of a group decision-making process.

- *Section 2.2: Applying Machine Learning Classifiers in Argumentation Context [20]*
  In this paper, it was proposed the use of machine learning classifiers to classify the direction (relation) between two arguments in the context of group decision-making. The use of machine learning techniques in the context of argumentation is described in this work, which involves negotiating arguments for and against a certain point of view. This process, known as Argument Mining, involves extracting arguments from unstructured texts and classifying the relations between them. Using machine learning classifiers to automatically identify the direction of an argument in a discussion makes it possible to improve the efficiency of GDSS. It was conducted an experimental evaluation of the approach using a dataset of annotated argumentation pairs. The results showed that the method achieved good precision, recall, and F1-score performance. The limitations of the approach and directions for future work were also discussed.

- *Section 2.3: Aspect Based Sentiment Analysis Annotation Methodology for Group Decision Making Problems: An Insight on the Baseball Domain [7]*
  This work focuses on Aspect-based sentiment analysis, a technique used to analyze and extract

the sentiment expressed towards specific aspects or features of an entity, such as a product or service. This technique can be useful in group decision-making problems, as it allows for a more nuanced and detailed understanding of the sentiments of different stakeholders. This paper presents a methodology for annotating text for aspect-based sentiment analysis in the context of group decision-making problems in the baseball domain. The paper describes the process of selecting and defining aspects for annotation and the steps involved in annotating text for sentiment toward these aspects. Besides that, an overview of the challenges and limitations of aspect-based sentiment analysis in baseball and discuss potential applications and future directions for this approach. Overall, this paper provides a valuable contribution to the field of sentiment analysis by presenting a methodology for annotating text for aspect-based sentiment analysis in the context of group decision-making problems, with a case study on the baseball domain.

- *Section 2.4: Supporting argumentation dialogues in Group Decision Support Systems: an approach based on dynamic clustering [23]*
  This paper proposes a method for supporting argumentation dialogues using dynamic clustering in group decision support systems. The method uses an unsupervised technique to group arguments into clusters based on discussed topics or alternatives. Experiments have been run with different combinations of word embedding, dimensionality reduction techniques, and clustering algorithms to determine the best approach. Conclusions showed that the KMeans++ clustering technique with SBERT word embedding and Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction resulted in the best performance. This method aims to improve the efficiency and effectiveness of argumentation dialogues in GDSS by providing a way to organize and group the arguments being made. This can ultimately lead to better decision-making by the group.

Chapter 3: Conclusions

The last chapter, Chapter 3, presents the resulting contributions from the works described in this thesis and also how they validate the hypotheses previously defined in the chapter 1. The activities developed during the period of this Ph.D. are also described in the remainder of this chapter. Section 3.2 presents other publications (subsection 3.2.2) that were made during this period related to other scientific projects but that are somehow related to the core topic of this thesis; collaboration and participation in scientific reviews and events are described in subsection 3.2.3; students supervision and the work they developed are presented in subsection 3.2.5. Finally, in section 3.3 some final remarks and considerations about the developed work are presented, as well as some guidelines for future work.

Chapter 2

Publications Composing the Doctoral Thesis

This chapter presents the set of four publications that have been selected to integrate this thesis and that describe the work carried out during the Ph.D. All publications are preceded by a table that presents all the information regarding them. Bellow, it presented the list of papers:

- **A web-based group decision support system for multi-criteria problems**
  Luís Conceição, Diogo Martinho, Rui Andrade, João Carneiro, Constantino Martins, Goreti Marreiros, and Paulo Novais. "A web-based group decision support system for multicriteria problems". In: *Concurrency and Computation: Practice and Experience* 33.2 (2021), e5298

- **Applying Machine Learning Classifiers in Argumentation Context**
  Luís Conceição, João Carneiro, Goreti Marreiros, and Paulo Novais. "Applying Machine Learning Classifiers in Argumentation Context". In: *International Symposium on Distributed Computing and Artificial Intelligence*. Springer. 2020, pp. 314–320

- **Aspect Based Sentiment Analysis Annotation Methodology for Group Decision Making Problems: An Insight on the Baseball Domain**
  Tiago Cardoso, Vasco Rodrigues, Luís Conceição, João Carneiro, Goreti Marreiros, and Paulo Novais. "Aspect Based Sentiment Analysis Annotation Methodology for Group Decision Making Problems: An Insight on the Baseball Domain". In: *World Conference on Information Systems and Technologies*. Springer. 2022, pp. 25–36

- **Supporting argumentation dialogues in Group Decision Support Systems: an approach based on dynamic clustering**
  Luís Conceição, Vasco Rodrigues, Jorge Meira, Goreti Marreiros, and Paulo Novais. "Supporting

Argumentation Dialogues in Group Decision Support Systems: An Approach Based on Dynamic Clustering". In: *Applied Sciences* 12.21 (2022), p. 10893

2.1 A web-based group decision support system for multi-criteria problems

Title	A web-based group decision support system for multi-criteria problems
Authors	Luís Conceição, Diogo Martinho, Rui Andrade, João Carneiro, Constantino Martins, Goreti Marreiros, and Paulo Novais
Publication Type	Journal
Publication Name	Concurrency and Computation Practice and Experience
Publisher	Wiley
Volume	33
Number	2
Pages	n/a
Year	2021
Month	January
Online ISSN	1532-0634
Print ISSN	1532-0626
URL	https://doi.org/10.1002/cpe.5298
State	Published
SJR impact factor (2021)	0.515, Computer Science Applications (Q2)
JCR impact factor (2021)	1.831, Computer Science, Theory & Methods (Q3)

A web-based group decision support system for multicriteria problems

Luís Conceição[1,2] | Diogo Martinho[2] | Rui Andrade[2] | João Carneiro[2] | Constantino Martins[2] | Goreti Marreiros[2] | Paulo Novais[1]

[1]Centro ALGORITMI, School of Engineering, University of Minho, Guimarães, Portugal
[2]Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD), Institute of Engineering, Polytechnic of Porto, Porto, Portugal

Correspondence
Luís Conceição, Centro ALGORITMI, School of Engineering, University of Minho, 4800-058 Guimarães, Portugal; or Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD), Institute of Engineering, Polytechnic of Porto, Porto 4200-072, Portugal.
Email: lmdsc@isep.ipp.pt

Funding information
FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology), Grant/Award Number: GrouPlanner Project (POCI-01-0145-FEDER-29178), Projects UID/CEC/00319/2019, and UID/EEA/00760/2019; Luís Conceição Ph.D., Grant/Award Number: SFRH/BD/137150/2018

Summary

One of the most important factors to determine the success of an organization is the quality of decisions made. Supporting a decision-making process is a complex task, mainly when decision-makers are dispersed. Group decision support systems (GDSSs) have been studied over the last decades with the goal of providing support to decision-makers; however, their acceptance by organizations has been difficult. This happens mostly due to usability problems, loss of interaction between decision-makers, and consequently, loss of information. In this work, we present a web-based GDSS developed to support groups of decision-makers, regardless of their geographic location. The system allows the creation of multicriteria problems and the configuration of the preferences, intentions, and interests of each decision-maker. The presented system uses a multiagent system to combine and process this information, using virtual agents that represent each decision-maker. We believe that, with this approach, we will proceed in the refinements of a successful GDSS to correctly support decision-makers while preserving the valuable intelligence and knowledge that can be generated in face-to-face meetings. Furthermore, the high level of usability that the system provides will contribute to an easier acceptance and adoption of this kind of systems.

KEYWORDS

automatic negotiation, group decision making, group decision support systems, multiagent systems

1 | INTRODUCTION

Group decision support systems (GDSSs) have been studied over the last decades with the goal of providing support to decision-makers that may be involved in group decision-making processes.[1-3] According to the literature, we know that decisions made in group can achieve better results compared with individual decisions. Furthermore, most of the organizations' organigrams require this type of decision-making process.[4] Nowadays, due to the actual paradigm of globalization, many companies are becoming global and turning into multinational organizations. As a result, managers (the decision-makers) spend most of their time traveling around the world and staying in different countries with different time zones, and become unavailable to gather at the same place and time to make a decision.

To overcome this issue, GDSS have been adapted, and we can now find many GDSS that are web-based to provide support to decision-makers in many areas of society, such as healthcare, economy, gastronomy, logistics, and industry. For example Miranda et al[5] proposed a simulated medical practice scenario to deal with staging cancer. In their proposal, the decisions were made in group and allowed collaborative work. These authors also implemented a multiagent system (MAS) to represent and exchange information related to real participants. In the work proposed by Tavana et al,[6] a GDSS was developed to evaluate and manage oil and natural gas transportations using the alternative pipeline routes from the Caspian Sea to other regions. They represented decision-makers' beliefs using the strength, weakness, opportunity and threat (SWOT) analysis with the Delphi method. These believes were integrated using the preference ranking organization method for enrichment evaluation

(PROMETHEE) to find the better solution for pipeline routes. In the work proposed by Morente-Molinera et al,[7], the authors developed a decision support system composed of both web and mobile applications to support the selection of wines. This system allowed decision-makers to participate in the decision-making process even if they were geographically dispersed. They considered the use of different techniques such as a fuzzy wine ontology, group decision support algorithms, and a fuzzyDL reasoner. Yazdani et al[8] proposed a GDSS for the selection of logistic providers. Their model combines a quality function deployment and the multicriteria decision analysis algorithm technique for order preference by the similarity to ideal solution (TOPSIS) to optimize a French logistic agricultural distribution center. They proposed a model that approaches the decision problem considering two perspectives: the technical and customer perspectives. To select third-party logistic providers, the system acts as an interface between decision-makers and customer values. To support agricultural parties, the system uses fuzzy linguistic variables in uncertain situations.

Regardless of the potential offered by web-based GDSS, success and acceptance of these systems by the decision-makers have not been positive so far. Some of the known reasons are related to the resistance to change from employees and the difficulties in the configuration of the system, either in the creation and configuration of the decision problem or in the configuration of the preferences for each decision-maker. Another reason that makes web-based GDSS hard to accept is related with the fear of losing dialog and idea discussion that can be achieved in face-to-face meetings.[9-11]

In this work, we present a web-based GDSS to support groups of decision-makers independently of their location. Our GDSS supports the group decision-making process for dispersed groups with users that cannot gather at the same place and time. The system allows the creation of multicriteria problems and the configuration of the preferences, intentions, and interests of each decision-maker. All the information gathered in each iteration is combined and processed in a MAS, which uses virtual agents that represent each decision-maker and act according to his/her preferences and intentions. These agents interact and negotiate with each other to find a solution for the selected problem with the goal of maximizing the group satisfaction regarding the proposed solution. We believe that the proposed GDSS can contribute to an increase in the acceptance of this type of systems by promoting the interaction between the members of the group, through the exchange of arguments regarding the alternatives and the criteria of the problem. On the other hand, the configurations of the preferences of the decision-makers in each iteration, can be easily configured through the interfaces developed to maximize the usability of the system.

The remaining sections of the paper are organized as follows. In Section 2, we present the proposed system, where we perform a general description of the GDSS, the concepts related to the system, and the description of the system's domain model. Section 3 presents a description of the GDSS workflow starting with the meeting creation and finishing with the report of the generated results. Finally, in Section 4, the conclusions are presented, along with some guidelines about future work.

2 | THE PROPOSED WEB-BASED GDSS

Our GDSS enables the group decision-making for multicriteria problems. The idea was to develop a system that could resemble a virtual meeting room but using the same logic applied in social networks. The user interface was built as a web application that enables all the interactions between the user and the system through any kind of device (such as a PC, a tablet, or a smartphone).

To better understand how the systems works, it is important to be aware of the concepts used in the GDSS.

- *Meeting* is a representation of a real group decision-making meeting in which one multicriteria problem will be discussed.
- *Problem* is the multicriteria problem, composed by a set of alternatives to solve that problem which are differentiated according to a set of criteria.
- *Topic* is a conversation topic that can be related with criteria or alternatives or both at the same time. Each one of the decision-makers can create topics and respond and evaluate the topics and messages created by other decision-makers. This conversation is related to either a public conversation (where each decision-maker can participate in the conversation) and to a private conversation (between two decision-makers, while exchanging requests, where only these two decision-makers can participate in the conversation). This way of exchanging information (using public and private conversation topics) has been inspired by the social networks logic and has been explored in a previous work.[12,13]
- *Decision-maker* is a person who participates in the group decision-making process. This person has access to the meeting information and can evaluate the multicriteria problem (by defining different preferences for each considered criterion and alternative) and may also define other personal configurations, such as the desired style of behavior, expectancy credibility, and expertise. All these concepts have been studied in previous studies and represent the intentions and goals of the decision-maker for the selected meeting.[14-17]
- *Style of behavior* is the expected behavior or the desired behavior of the agent in the negotiation process. We have followed the work and concepts proposed by Carneiro et al[14], and we have identified five main styles of behavior, which are integrating, compromising, dominating, avoiding, and obliging. These five styles are differentiated in four dimensions that represent how the decision-maker intends to behave throughout the decision-making process. These dimensions are the concern for self (importance given towards self-interests and goals), concern for others (importance given towards other decision-makers' interests and goals), activity level (the participation effort that is related

to the probability to create conversation topics), and resistance to change (the probability to accept or refuse incoming requests to change preferences).

- *Credibility* is defined in this work as the possibility for each decision-maker to select which other decision-makers he/she considers to be credible for the corresponding meeting. This selection is related to concepts such as trust, reliability accuracy, quality or even authority, reputation, and competence. We have explored this concept in more detail in a previous work.[18]
- Expectancy – in this work, we have defined expectancy as the perception that the decision-maker has regarding the acceptance of his preferences by other decision-makers. This expectation can influence satisfaction and may have a negative, positive, or neutral impact depending on whether the expectations are achieved, exceeded, or not achieved. We have explored this concept with more detail in a previous work.[19]
- *Satisfaction* is related to the perception of the quality of the decision. We have studied this concept in previous works,[19,20] and it can be measured according to the decision-maker's expectations, style of behavior, emotional changes, and mood variation.
- *Expertise* corresponds to the decision-maker's self-evaluation of his/her expertise level for the corresponding meeting. This concept has been studied in (including credibility and expertise), and we have identified five levels of expertise, which are expert, high, medium, low, and null.[18]
- *Available time* indicates the time needed for a decision-maker to analyze the problem. This corresponds to the availability specified by the decision-maker towards the decision-making process and whether he/she intends to receive detailed information by the system. If the available time is low, the information provided by the system should be more specific and oriented to the interests of the decision-maker, while if the available time is high, this information may not be only related to the decision-maker's self-interests but also related to the interests of other decision-makers. We have explored this concept with more detail in the work of Carneiro et al[21]

The system is composed by a MAS, and for each decision-making process, a group of agents will be used where each agent will act according to the decision-maker' preferences that it represents. Furthermore, each agent will try to obtain a solution for the multicriteria problem using an argumentation-based dialog model. Agents use deliberative dialogs tp identify the most consensual decision that brings the highest satisfaction to all decision-makers (which could correspond to the selection of one or more alternatives as a proposed solution for the problem). In addition to that, agents can also use other kind of dialogs, such as negotiation persuasion and information seeking.[16,17]

2.1 | Domain model

Figure 1 shows a high-level representation of the domain model of the application without any implementation details. The main concept in the application is obviously the meeting, everything else in the application, aside from user account management, revolves around the meeting. Since a group decision-making process is an iterative process, the system allows each meeting to have several iterations, where each iteration can have

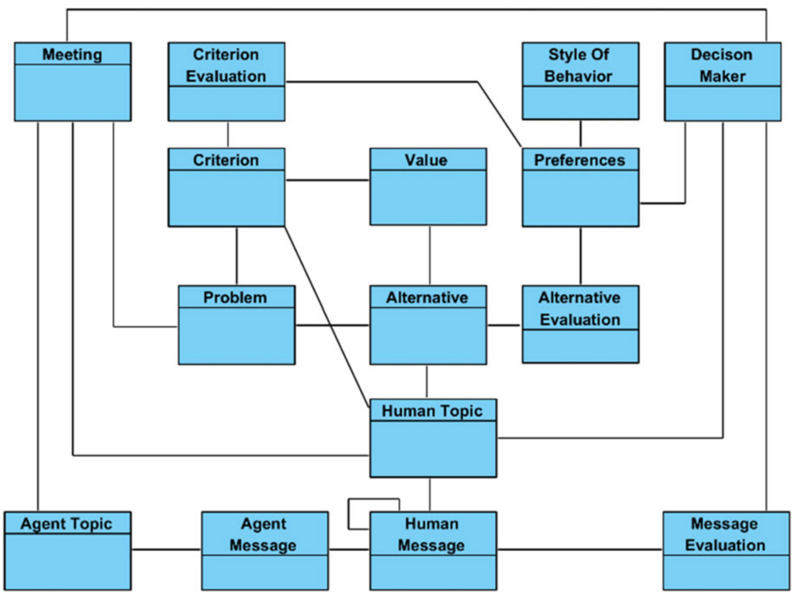

FIGURE 1 The proposed system's domain model

different problem configurations as well as different decision-makers. This way the system will be able to deal with situations where one or more decision-makers may abandon the decision-making process in the end of one iteration and before the meeting is concluded. Likewise, the system will be able to in include new decision-makers in the decision-making process if it is necessary.

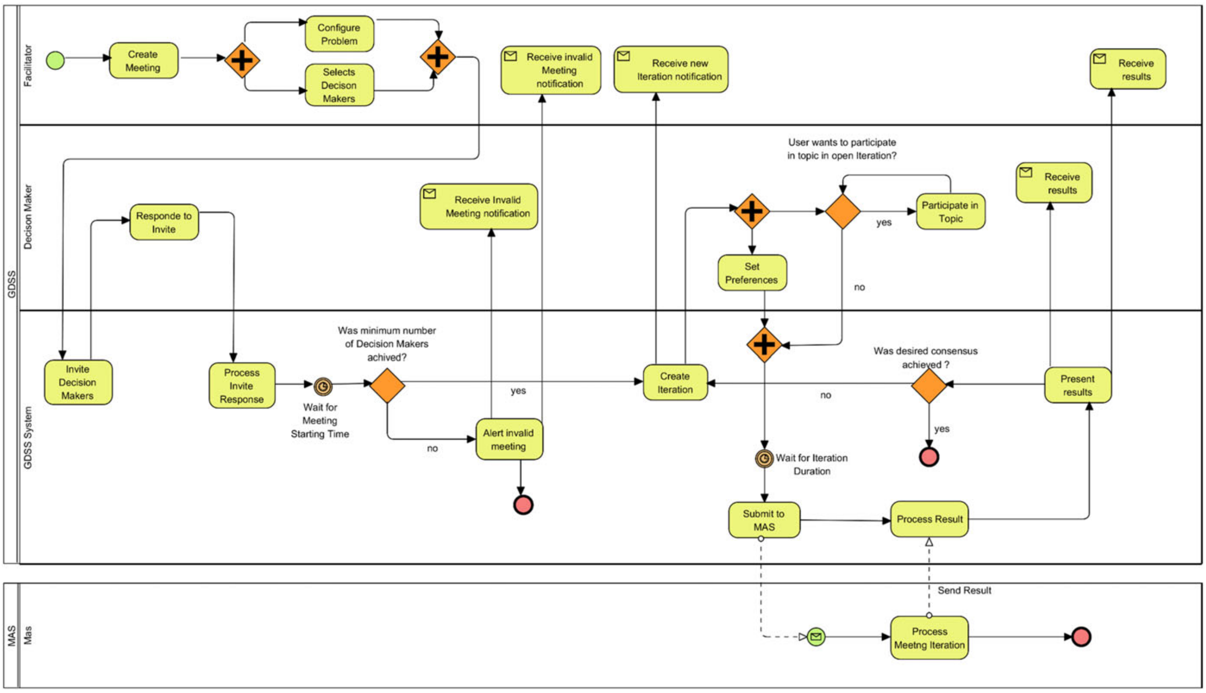

FIGURE 2 BPMN diagram of GDSS

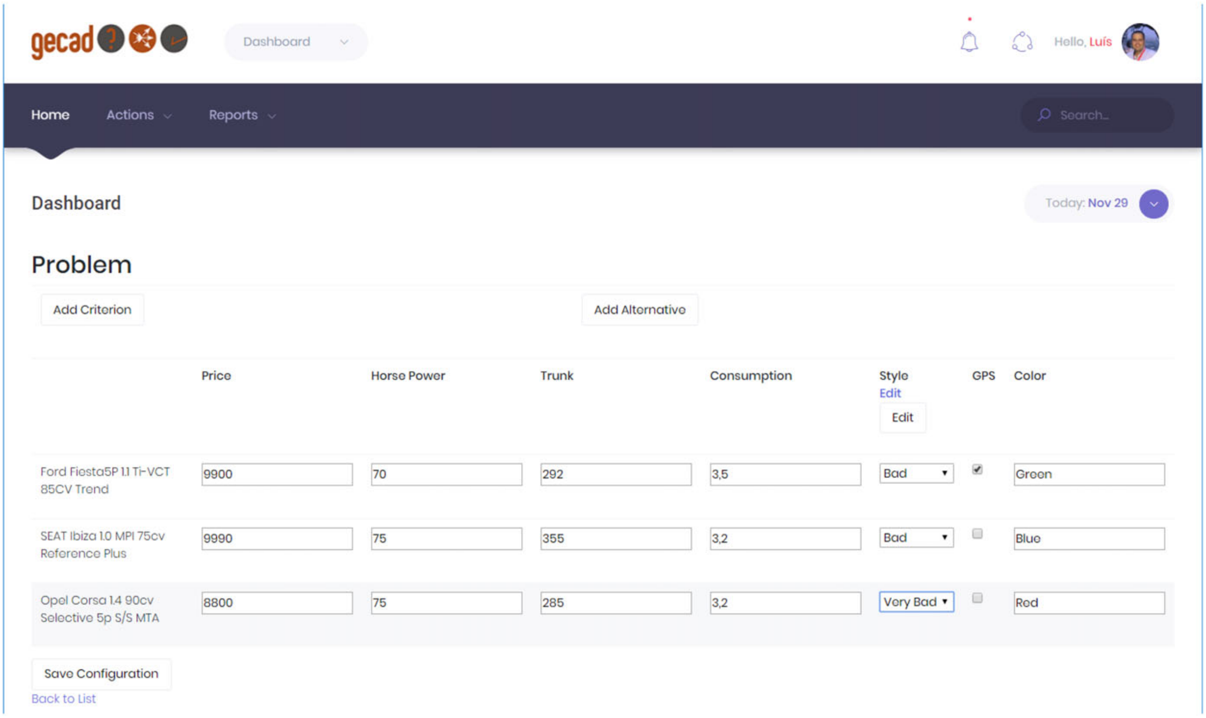

FIGURE 3 Problem configuration interface

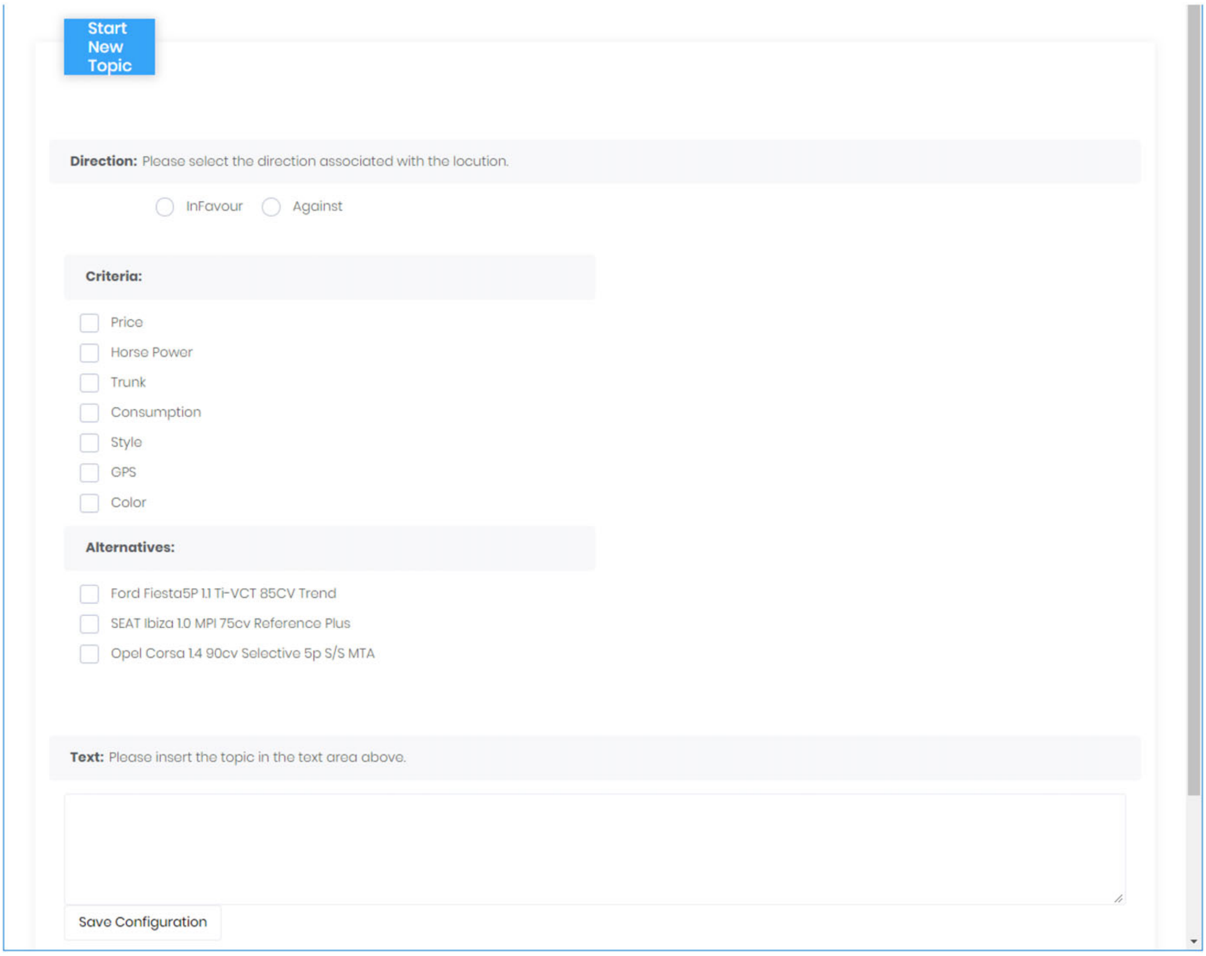

FIGURE 4 Topic creation interface

FIGURE 5 Respond to message interface

The meeting contains all the information related with each decision-making process, of which we highlight the following: a definition of the Problem, a group of decision-makers, a list of decision-makers' conversations, and a list of agents' conversations.

The problem model defines the multicriteria problem as a set of criteria and a set of alternatives. To complete the problem definition, it is also necessary to specify the relation between the criteria and the alternatives, which corresponds to the value that each alternative has in each criterion. It should be noticed however that the system can handle criteria from various types such as subjective, numeric, Boolean, and classificatory. In addition to that, greatness is associated to each criterion, which can be of maximization or minimization. For example, if we want to minimize the prize criterion, then the cheapest product would be the most beneficial solution.

As for the decision-maker, it will have his/her own preferences for each iteration. The preferences are in turn the decision-maker's expectancy regarding the selected alternative, the style of behavior, the decision-maker's expertise level, the available time, and credible decision-makers. Finally, preferences may also include the decision-makers' evaluation for the alternatives and criteria.

The human topic represents a conversation between decision-makers about one or more criteria and/or one or more alternatives. A human topic is created by a decision-maker and contains the initial human message for that human topic. As for human messages, they will either be an opening message or responses in a topic. Whenever a human message is included in the topic, it will be associated with the message it is responding to, as well as a message evaluation to that message.

Finally, we have agent topics, which represent the dialogs created by the MAS (messages exchanged between agents). If an agent message is derived from a human message, the agent message will inherit the characteristics of the corresponding human message.

The MAS used in this GDSS uses a framework that encapsulates the JADE framework with the intention of representing or virtualizing the interaction between decision-makers in face-to-face meetings, allowing the implementation of different dialog models and agents' behaviors.[22] This framework implements a type of communication between agents that guarantees that, at any given moment of time, all agents are in possession of the same knowledge and are therefore capable of simulating what could happen in a face-to-face meeting (in this case, whenever a decision-maker decides on a subject, all participants receive this information at the same instant of time). This approach uses a social network logic in which conversations are maintained in the form of topics where each agent creates a new topic for each subject, and then, other agents in the group can argue regarding that topic. The process ends whenever all involved agents withdraw from the discussion, which corresponds to them not wanting to discuss new topics nor responding to existing topics.

3 | WORKFLOW OF A GROUP DECISION-MAKING PROCESS

Assuring usability was a mandatory requirement throughout the development process of this GDSS, to simplify as possible the use of the system by a decision-maker. To better describe the workflow of this application, we used a business process model and notation (BPMN) diagram that is represented in Figure 2.

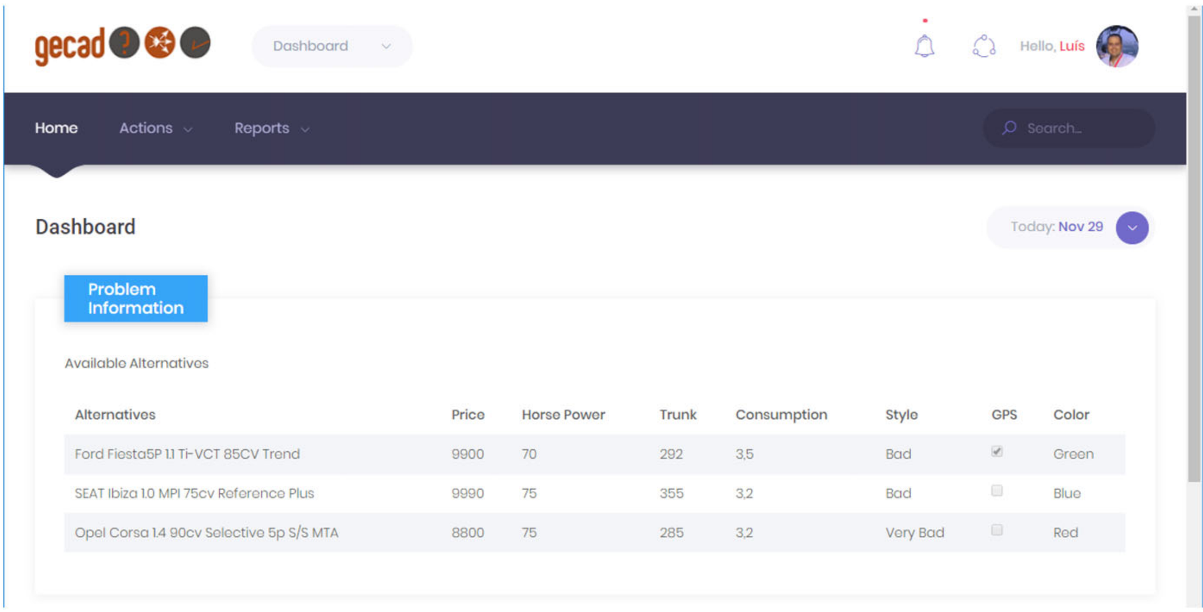

FIGURE 6 Decision-maker preference configuration (problem information) interface

The process begins when a user creates a meeting and consequentially becomes the meeting facilitator for the newly created meeting. The facilitator is then responsible for making the problem configuration, namely the specification of criteria and alternatives, as shown in Figure 3. It is important to notice that the facilitator can easily add new criteria or alternatives to the problem matrix simply by clicking in the "Add Criterion" or "Add Alternative" buttons available on the top of the table. In addition to that, facilitator also needs to invite other users to participate as decision-makers in the meeting. The invited users will receive a notification requesting their participation in the meeting and will then be able to accept or decline the invitation.

After each intended user is invited to participate in the meeting, the system will wait until the meeting starting time is ready. The system will then verify if the minimum amount of required decision-makers for the meeting was achieved, and in this case, the process will proceed to the iterative decision phase. Otherwise, if the number of decision-makers is below the minimum, then the process will terminate, and the system will notify the facilitator and the invited users that the meeting was invalid.

When starting an iteration, both the facilitator and the available decision-makers will be notified. From this moment on, all decision-makers can create discussion topics regarding alternatives or criteria (as shown in Figure 4). To create a new topic, the decision-maker must first select the direction associated with the locution, then select the criteria and/or alternatives that are related with the topic, and finally write the locution to conclude the topic creation. When a new topic is created, the system will notify all the available decision-makers who are participating in the decision-making process. After that, all the available decision-makers can respond to the topics and assess their importance.

As shown in Figure 5, a response message indicates the message that a user is responding to and its direction (if the response is in favor or against that message), as well as if that message is related with the criteria, alternatives, or both. In the response, the decision-maker needs

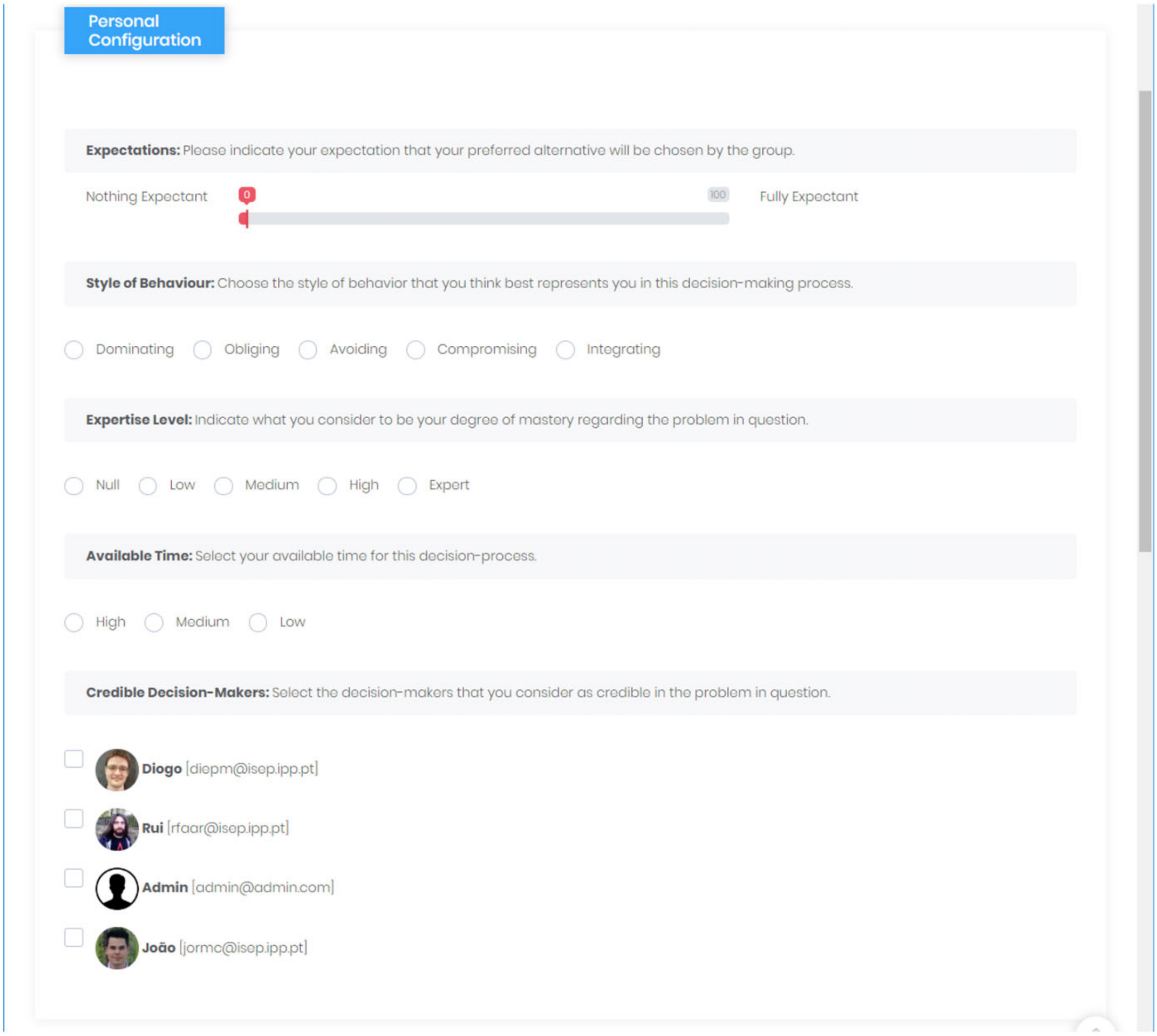

FIGURE 7 Decision-maker preference configuration (personal configuration) interface

to evaluate the message. There are three possible outcomes associated to this evaluation, the first being related to an evaluation greater than zero (in this case, the response will be a reinforcement to the original message, and then he can write a response reinforcing that message). The second case is related to an evaluation lower than zero (and in this case, the response will be an attack to the original message). The third and final outcome is related to evaluations equal to zero, and in this case, the system will not allow the introduction of a response message because it is considered that the decision-maker does not have an opinion about the original message.

Alongside the creation of discussion topics and responses to messages, decision-makers must also configure their preferences. The preference configuration interface was designed according to a template that was developed specifically for a web-based GDSS and that demonstrated high usability and configuration speed for the decision-makers.[11] This interface is composed of three main sections: problem information, personal configuration, and problem configuration. The problem information section presents the multicriteria problem to the decision-maker allowing the

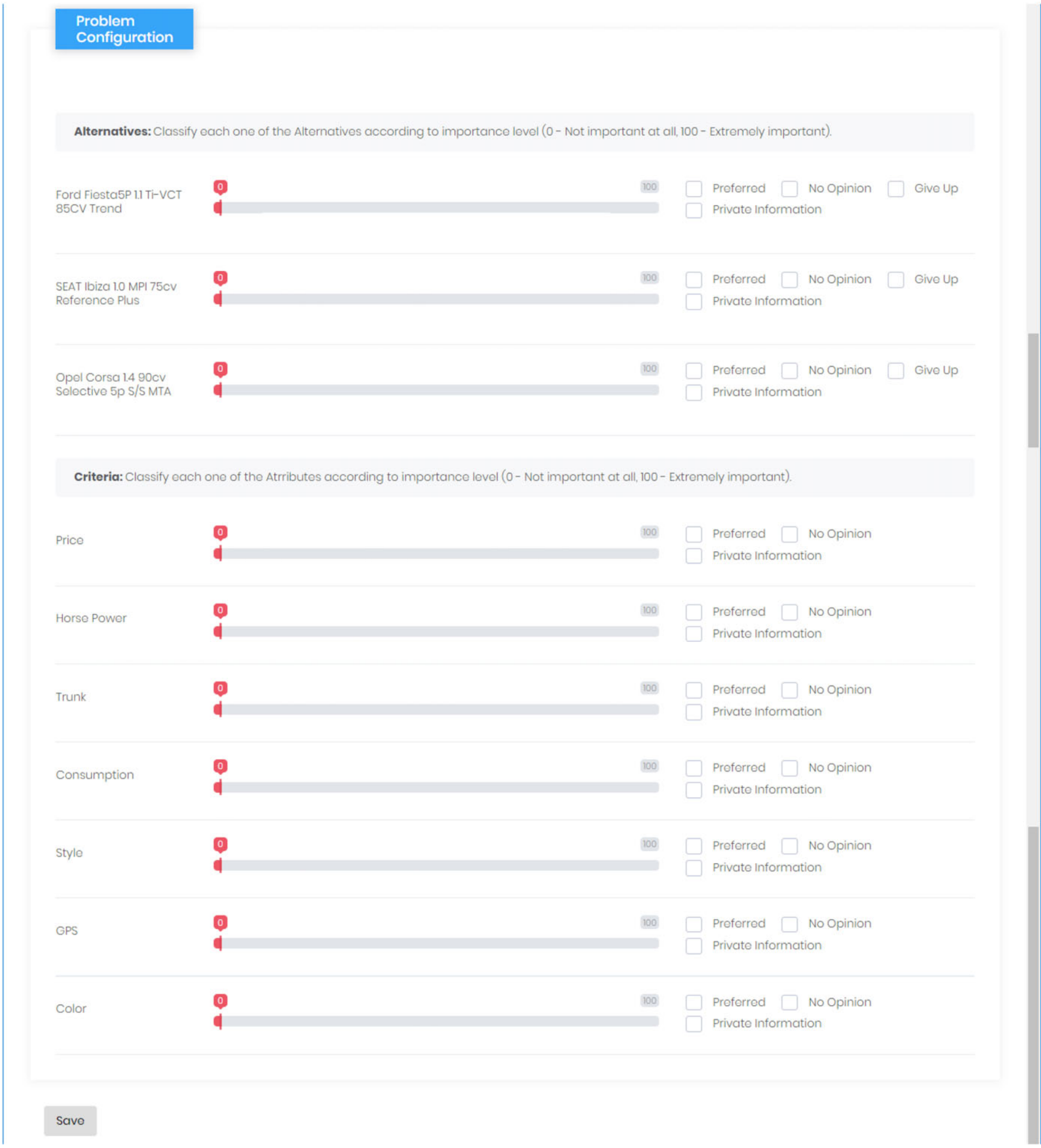

FIGURE 8 Decision-maker preference configuration (problem configuration) interface

analysis of the alternative's values for each criterion (see Figure 6). In the personal configuration section (see Figure 7), the decision-maker needs to indicate its expectations regarding his preferred alternative to be the one chosen by the group, the desired style of behavior for the agent that will represent him in the negotiation process, the level of expertise concerning the subject of the decision problem, the available time to spend in the process (analyzing results, etc), and finally the decision-maker must also indicate which decision-makers it deems credible among the others regarding the problem being discussed. The last section of the interface is the problem configuration section that is related with the evaluation of criteria and alternatives (see Figure 8). This evaluation is done in a range between 0 and 100. To evaluate each one of the criteria and alternatives, the decision-maker can easily slide or click on the slide bar. By presenting all the evaluations together on top of one another while the user is performing the evaluations, it will allow him/her to easily compare each evaluation and assign his/her preferences more accurately.

After each decision-maker provides their preferences and configurations, the system will wait again until the iteration is completed. After this, the system will send all the meeting data to the MAS. At this point, the MAS will start the negotiation process with the received data. When the process ends, the MAS will send back the results to the GDSS. These results include all the messages exchanged between the agents during the negotiation process, as well as the achieved consensus with the results of the negotiation process and the satisfaction level measured for each one of the decision-makers regarding the selection of each one of the alternatives.

After all the results are generated and received from the MAS, the system will notify the decision-makers to consult them and will also verify if the desired level of consensus was achieved. If it was achieved, then the iterative decision process will finish, and the system will notify the facilitator and the decision-makers with the final meeting results. Otherwise, the iterative decision process will continue, and each decision-makers will be able to review the current results and reconfigure his/her initial preferences. The conditions mentioned previously will be verified again to start a new iteration, and this process will be repeated until a consensus is achieved.

Decision-makers can access the results of each iteration through a dashboard that presents intelligent reports. In this dashboard, the information presented is directed to the decision-maker and can vary according to three factors: level of expertise, available time, and the level of interest in the process.[21,23] Figure 9 presents a dashboard with the results of an iteration. The results are presented in two main sections. The first section presents statistical data where the decision-maker can analyze the support of alternatives and criteria, as well as the most consensual alternative at the end of the iteration, his/her satisfaction regarding the selected alternative, and the group satisfaction with the selection of that alternative. In addition to that, in the left chart, the decision-maker can observe the support of each one of the alternatives and the corresponding group

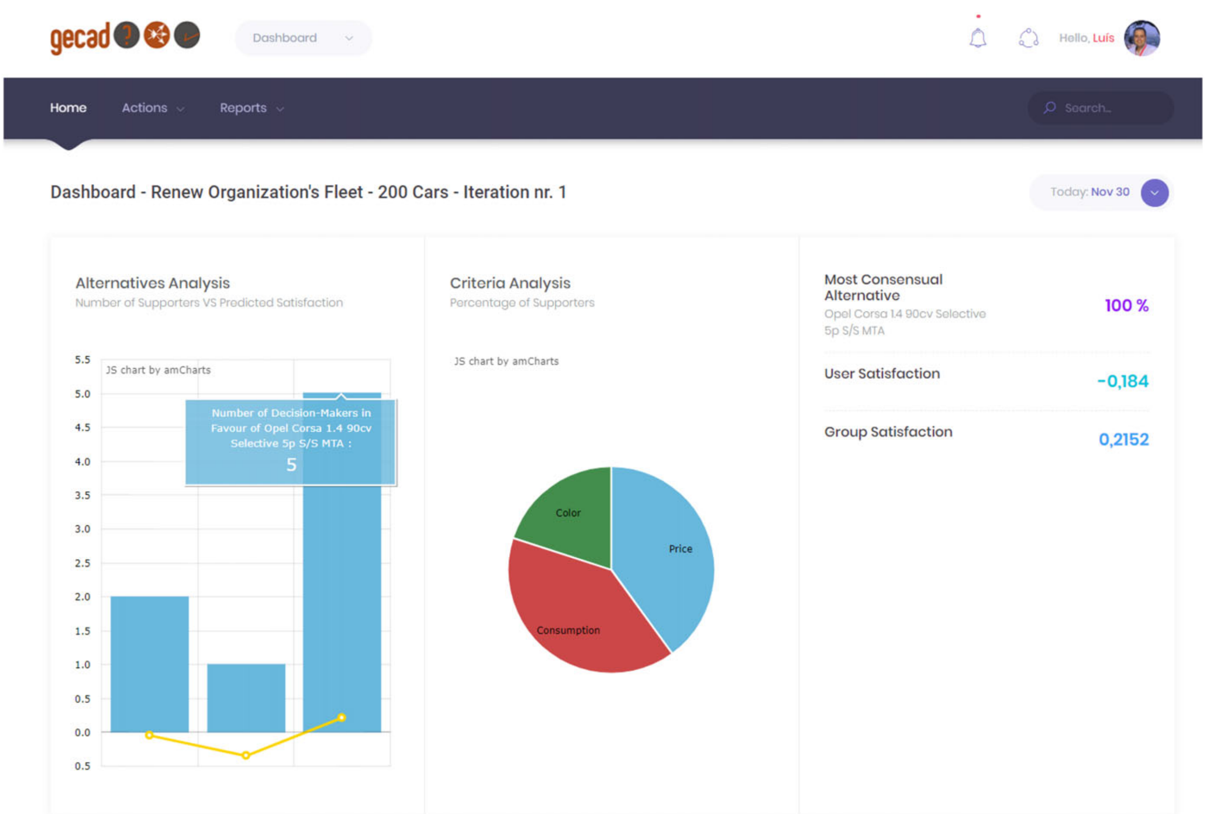

FIGURE 9 Iteration results report (Part 1)

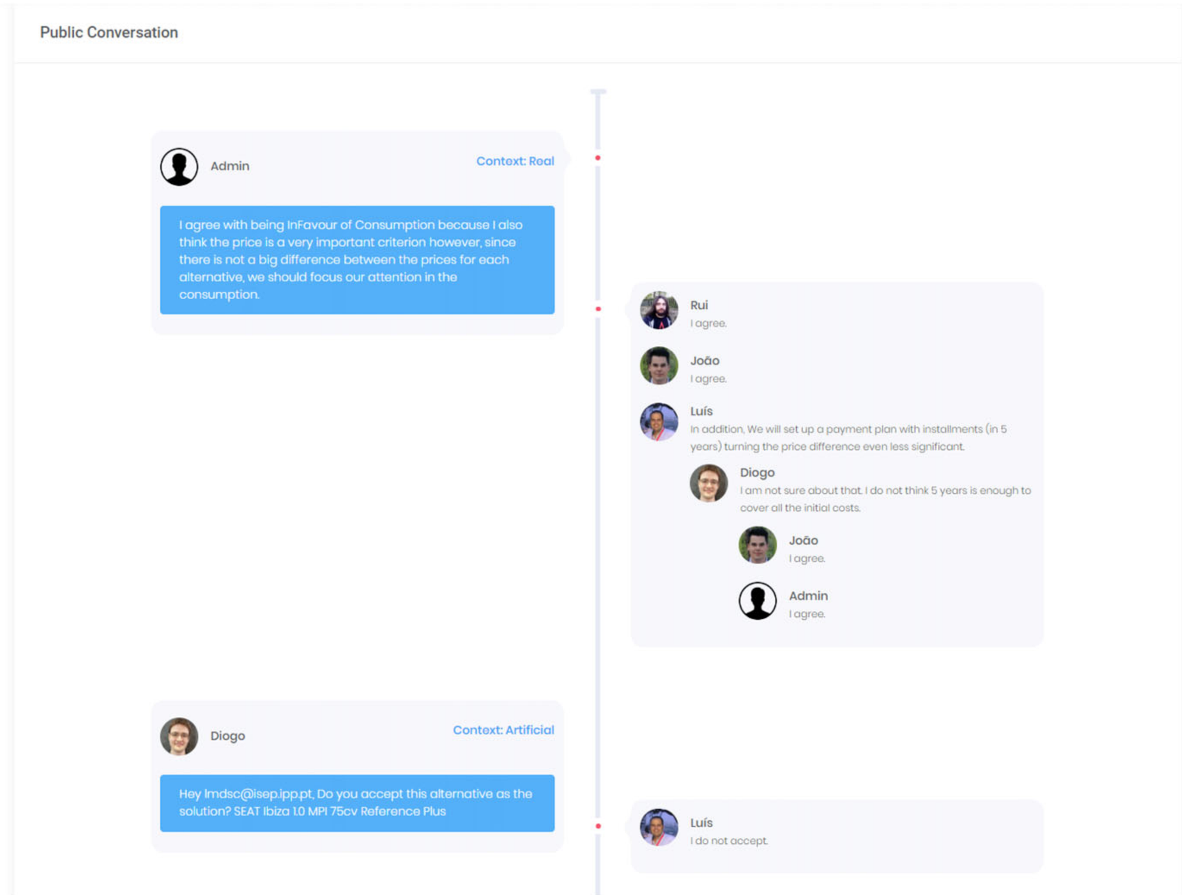

FIGURE 10 Iteration results report (Part 2)

satisfaction (yellow line) in case each one of those alternatives were to be selected as the decision for that iteration. The pie chart indicates the support towards each considered criterion.

The second section (see Figure 10) presents nonstatistical data, which include all the messages exchanged during the negotiation process performed by the MAS for that iteration. The MAS is able to use and understand the topics created by the decision-makers, which correspond to real-context messages, and generates new messages (such as requests) that corresponds to an artificial-context message.

4 | CONCLUSIONS AND FUTURE WORK

In a world that increasingly becoming more global, we are now observing remarkable changes in today's society and in many different traditional and conventional processes such as the decision-making process. What was once a more individualistic process which then evolved into a group decision-making process is now outdated due to arising constraints of this globalization. It no longer makes sense to gather decision-makers at the same time and place to make decisions, and the process must evolve to support decision-makers spread around the world, staying in different countries with different time zones. As a result, we are now dealing with a new type of decision support systems, also known as web-based GDSS.

A lot of work and efforts should be taken before web-based GDSS are accepted, mostly related with the resistance to change and the capability to correctly model the intentions and preferences of the decision-maker while preserving the advantages that are inherent to face-to-face meetings. In this work, we deal with these aspects, and we have presented a web-based GDDS to support the group decision-making process for dispersed groups with users that cannot gather at the same place and at the same time. The system allows the creation of multicriteria problems and the configuration of the preferences, intentions, and interests of each decision-maker. The system makes use of a MAS to combine and process this information, using virtual agents that represent each decision-maker and act according to his/her configurations. These agents interact and negotiate with each other to find a solution for the selected problem.

The GDSSs that we referenced in Section 1 were mostly developed to support the decision-making of a specific problem; in this work, the proposed GDSS allows the configuration of any multicriteria problem, namely its alternatives and the criteria that make it possible to value each of the alternatives. We believe that with this approach, we will proceed in the refinements of a successful GDSS to correctly support decision-makers while preserving the valuable intelligence and knowledge that can be generated in face-to-face meetings. Furthermore, the high level of usability that the system provides will contribute to an easier acceptance and adoption of this kind of systems.

For future work, we aim to study and develop models using machine learning techniques to extract knowledge from argumentative-based dialog models performed by both decision-makers and agents. IN particular, it is intended to model argumentative processes in GDSS, using MASs, considering the decision-makers' objectives and understanding the decision process. Furthermore, with this work, we intend to create models to analyze, classify, and process these data to potentiate the generation of new knowledge that will be used by both agents and decision-makers.

ACKNOWLEDGMENTS

This work was supported by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) with GrouPlanner Project (POCI-01-0145-FEDER-29178) and within the Projects UID/CEC/00319/2019 and UID/EEA/00760/2019, and the Luís Conceição Ph.D. Grant with the reference SFRH/BD/137150/2018.

ORCID

Luís Conceição https://orcid.org/0000-0003-3454-4615
Diogo Martinho https://orcid.org/0000-0003-1683-4950
João Carneiro https://orcid.org/0000-0003-1430-5465
Constantino Martins https://orcid.org/0000-0001-9344-9832
Goreti Marreiros https://orcid.org/0000-0003-4417-8401
Paulo Novais https://orcid.org/0000-0002-3549-0754

REFERENCES

1. DeSanctis G, Gallupe B. Group decision support systems: a new frontier. *Database Adv Inf Syst.* 1985;16:3-10.
2. DeSanctis G, Gallupe B. A foundation for the study of group decision support systems. *Manage Sci.* 1987;33:589-609.
3. Gallupe B, DeSanctis G, Dickson GW. The impact of computer-based support on the process and outcomes of group decision making. Paper presented at: 7th International Conference on Information Systems; 1986; San Diego, CA.
4. Luthans F. *Organizational Behavior: An Evidence-based Approach.* New York, NY: McGraw-Hill Irwin; 2011.
5. Miranda M, Abelha A, Santos M, Machado J, Neves J. A group decision support system for staging of cancer. Paper presented at: First International Conference on Electronic Healthcare; 2008; London, UK
6. Tavana M, Behzadian M, Pirdashti M, Pirdashti H. A PROMETHEE-GDSS for oil and gas pipeline planning in the Caspian Sea basin. *Energy Econ.* 2013;36:716-728.
7. Morente-Molinera JA, Wikström R, Herrera-Viedma E, Carlsson C. A linguistic mobile decision support system based on fuzzy ontology to facilitate knowledge mobilization. *Decis Support Syst.* 2016;81:66-75.
8. Yazdani M, Zarate P, Coulibaly A, Zavadskas EK. A group decision making support system in logistics and supply chain management. *Expert Systems with Applications.* 2017;88:376-392.
9. van Hillegersberg J, Koenen S. Adoption of web-based group decision support systems: experiences from the field and future developments. *Int J Inf Syst Proj Manag*. 2016;4:49-64.
10. van Hillegersberg J, Koenen S. Adoption of web-based group decision support systems: conditions for growth. *Procedia Technol.* 2014;16:675-683.
11. Carneiro J, Martinho D, Marreiros G, Novais P. A general template to configure multi-criteria problems in ubiquitous GDSS. *Int J Software Eng Appl.* 2015;9:193-206.
12. Carneiro J, Marreiros G, Novais P. An approach for a negotiation model inspired on social networks. Paper presented at: International Workshops of PAAMS 2015; 2015; Salamanca, Spain.
13. Carneiro J, Martinho D, Marreiros G, Novais P. Intelligent negotiation model for ubiquitous group decision scenarios. *Front Inf Technol Electron Eng.* 2016;17:296-308.
14. Carneiro J, Saraiva P, Martinho D, Marreiros G, Novais P. Representing decision-makers using styles of behavior: an approach designed for group decision support systems. *Cogn Syst Res.* 2018;47:109-132.
15. Carneiro J, Conceição L, Martinho D, Marreiros G, Novais P. Including cognitive aspects in multiple criteria decision analysis. *Ann Oper Res.* 2016;1(23):
16. Carneiro J, Martinho D, Marreiros G, Jimenez A, Novais P. Dynamic argumentation in UbiGDSS. *Knowl Inf Syst.* 2018;55:633-669.
17. Carneiro J, Martinho D, Marreiros G, Novais P. Arguing with behavior influence: a model for web-based group decision support systems. *Int J Inf Tech Decis.* 2019;18(2):517-553.
18. Carneiro J, Martinho D, Marreiros G, Novais P. Including credibility and expertise in group decision-making process: an approach designed for UbiGDSS. Paper presented at: World Conference on Information Systems and Technologies; 2017; Porto Santo Island, Portugal.
19. Carneiro J, Santos R, Marreiros G, Novais P. Evaluating the perception of the decision quality in web-based group decision support systems: a theory of satisfaction. Paper presented at: International Workshops of PAAMS 2017; 2017; Porto, Portugal.
20. Carneiro J, Marreiros G, Novais P. Using satisfaction analysis to predict decision quality. *Int J Artif Intell.* 2015;13:45-57.

21. Carneiro J, Conceição L, Martinho D, Marreiros G, Novais P. Intelligent reports for group decision support systems. In: Novais P, Konomi S, eds. *Intelligent Environments 2016: Workshop Proceedings of the 12th International Conference on Intelligent Environments*. Amsterdam, The Netherlands: IOS Press; 2016.
22. Carneiro J, Alves P, Marreiros G, Novais P. A multi-agent system framework for dialogue games in the group decision-making context. Paper presented at: WorldCIST'19-7th World Conference on Information Systems and Technologies; 2019; Galicia, Spain.
23. Conceição L, Carneiro J, Martinho D, Marreiros G, Novais P. Generation of intelligent reports for ubiquitous group decision support systems. Paper presented at: 2016 Global Information Infrastructure and Networking Symposium; 2016; Porto, Portugal.

How to cite this article: Conceição L, Martinho D, Andrade R, et al. A web-based group decision support system for multicriteria problems. *Concurrency Computat Pract Exper*. 2019;e5298. https://doi.org/10.1002/cpe.5298

2.2 Applying Machine Learning Classifiers in Argumentation Context

Title	Applying Machine Learning Classifiers in Argumentation Context
Authors	Luís Conceição, João Carneiro, Goreti Marreiros, and Paulo Novais
Publication Type	Conference Proceedings
Publication Name	Advances in Intelligent Systems and Computing
Publisher	Springer, Cham
Volume	1237
Pages	314-320
Year	2020
Month	August
Online ISBN	978-3-030-53036-5
Print ISBN	978-3-030-53035-8
URL	https://link.springer.com/chapter/10.1007/978-3-030-53036-5_34
State	Published

Applying Machine Learning Classifiers in Argumentation Context

Luís Conceição[1,2 [0000-0003-3454-4615]], João Carneiro[2[0000-0003-1430-5465]], Goreti Marreiros[2[0000-0003-4417-8401]] and Paulo Novais[1[0000-0002-3549-0754]]

[1] ALGORITMI Centre, University of Minho, Guimarães 4800-058, Portugal
[2] GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Institute of Engineering, Polytechnic of Porto, 4200-072 Porto, Portugal
{msc,jrc,mgt}@isep.ipp.pt
pjon@di.uminho.pt

Abstract. Group decision making is an area that has been studied over the years. Group Decision Support Systems emerged with the aim of supporting decision makers in group decision-making processes. In order to properly support decision-makers these days, it is essential that GDSS provide mechanisms to properly support decision-makers. The application of Machine Learning techniques in the context of argumentation has grown over the past few years. Arguing includes negotiating arguments for and against a certain point of view. From political debates to social media posts, ideas are discussed in the form of an exchange of arguments. During the last years, the automatic detection of this arguments has been studied and it's called Argument Mining. Recent advances in this field of research have shown that it is possible to extract arguments from unstructured texts and classifying the relations between them.
In this work, we used machine learning classifiers to automatically classify the direction (relation) between two arguments.

Keywords: Argument Mining, Machine Learning Classifiers, Argumentation-based dialogues.

1 Introduction

Group decision making is an area that has been studied over the years. In organizations, most decisions are taken in groups, either for reasons of organizational structure, as their organizational organigrams oblige them to do so, or for the associated benefits of group decision making, such as: sharing responsibilities, greater consideration of problems and possible solutions, and also allow less experienced elements learn during the process [1-3].

Group Decision Support Systems (GDSS) emerged with the aim of supporting decision makers in group decision-making processes. They have been studied over the past 30 years and have become a very relevant research topic in the field of Artificial Intelligence, being nowadays essentially developed with web-based interfaces [4-8].

The globalization of markets and the appearance of large multinational companies mean that many of the managers are constantly traveling through different locations and in areas with different time zones [9].

In order to properly support decision-makers these days, it is essential that GDSS provide mechanisms such as automatic negotiation, represent interests of decision-makers, potentiate the generation of ideas, allow the existence of a process, between others.

Our research group has been working over the last few years on methods and tools that support decision-makers that are geographically dispersed, more specifically on argumentation-based frameworks [10]; in the definition of behaviour styles for agents that represent real decision-makers [11, 12]; in the satisfaction of the decision-makers regarding the group decision-making process [13], as well as in the group's satisfaction in relation to the decision taken.

Our dynamic argumentation framework allows GDSS to be equipped with features that allow decision-makers, even when they are geographically dispersed, to benefit from the typical advantages associated to the face-to-face group decision-making processes. This framework accompanies the decision-maker throughout the group decision-making process, through the implementation of a multi-agent system, where each agent represents a human decision-maker in the search for a solution for the problem, proposing one or more alternatives as solution to the problem from the set of initial alternatives, taking into account the preferences and interests of the decision-maker in the decision problem that is being handled. The dynamic argumentation framework allows decision-makers to understand the conversation performed between agents in the negotiation process, but also the agents (**Fig. 1**) are able to understand the new arguments introduced and exchanged by decision-makers. Agents can use these new arguments to advice decision-makers and to find solutions during the decision-making process [10].

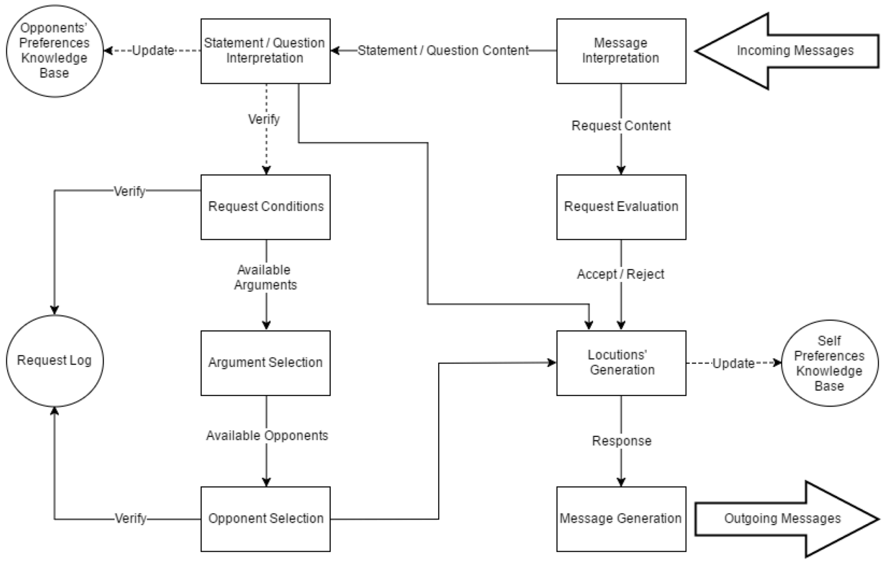

Fig. 1. Agent's Communications Workflow [11]

The application of Machine Learning (ML) techniques in the context of argumentation has grown over the past few years. Argument Mining (AM) is the automatic identification and extraction of the structure of inference and reasoning expressed in the form of arguments in natural language [14].

In this work our goal is to automatically classify the relation between two arguments introduced by decision-makers in our dynamic argumentation framework. The framework works as a social network where each decision-maker can make a post expressing his/her opinion regarding one (or more) alternative(s) and/or criteria, and the decision-maker has to classify the direction of the argument between "against" (if his/her comment is attacking the idea that he/she is responding) or "in favour" (if his/her comment is supporting the idea that he/she is responding). To create a model to automatically classify the relations between arguments we used the dataset created by Benlamine, Chaouachi, Villata, Cabrio, Frasson and Gandon [15] which consists of a set of arguments 526 arguments exchanged by participants in online debates regarding different topics. The dataset is annotated with the relation between each two arguments in support (if the argument is in favour of the idea expressed in the previous argument) and attack (if the argument is against the idea expressed in the previous argument). With this dataset we ran experiments to train a classifier for the relation between arguments.

The paper is organized as follows. Section 2 we present some of the most relevant works in the literature about AM. Section 3 describes our work and experiences. In

Section 4 we discuss the obtained results of the trained classifier. Finally, in Section 5 we describe the conclusions and some ideas about future work that we intend to follow.

2 Related Work

It is possible to find different approaches in this area, for instance in the work of Rosenfeld and Kraus [16] they used ML techniques in a multi-agent system that supports humans in argumentative discussions by proposing possible arguments to use. To do this, they analysed the argumentative behaviour from 1000 human participants and they proved that is possible to predict human argumentative behaviour using ML techniques [16].

Another interesting work is the one published by Carstens and Toni [17] where they worked in the identification and extraction of argumentative relations. To do that, they classify pairs pf sentences according to whether they stand in an argumentative relation to other sentences, considering any sentence as argumentative that supports or attacks another sentence.

In Cocarascu and Toni [18] they propose a deep learning architecture to identify argumentative relations of attack and support from one sentence to another, using Long Short-Term Memory (LSTM) networks. Authors concluded that the results improved considerably the existing techniques for AM.

In Rosenfeld and Kraus [19] they presented a novel methodology for automated agents for human persuasion using argumentative dialogs. This methodology is based on an argumentation framework called Weighted Bipolar Argumentation Framework (WBAF) and combines theoretical argumentation modelling, ML and Markovian optimization techniques. They performed field experiments and concluded that their agent is able to persuade people no worse than people are able to persuade each other.

Swanson, Ecker and Walker [20] published a paper where they created a dataset by extracting dialogues from online forums and debates. They established two goals: the extraction of arguments from online dialogues and the identification of argument's facet similarity. The first consists in the identification and selection of text excerpts that contain arguments about an idea or topic, and the second is about argument's semantic, an attempt to identify if two arguments have the same meaning about the topic.

In Mayer, Cabrio, Lippi, Torroni and Villata [21] authors performed argument mining in clinical trials in order to support physicians in decision making processes. They extracted argumentative information such as evidence and claims, from clinical trials which are extensive documents written in natural language, using the MARGOT system [22].

3 Training the Classifier of Arguments Relations

In this Section we describe the process we have followed to train our argument relation classifier. As we stated in Section 1, the dataset used in this work consists in a set of arguments extracted by Benlamine, Chaouachi, Villata, Cabrio, Frasson and Gandon

[15] from 12 different online debates. This dataset consists in a set of 526 arguments, composing 263 pairs of arguments extracted from the 12 online debates about different topics such as abortion, cannabis consumption, bullism, among others. We extracted the data from the xml files and built the initial dataset (represented by an excerpt in Table 1) with 3 columns named: “arg1”, “arg2”, and “relation”, corresponding the first to the first argument, the second to the argument that responds to the first, and the relation which consists in the annotation that characterizes the relation between the arguments (“InFavour”/”Against”). The annotated dataset contains 48% (127) argument pairs classified as “InFavour” and 52% (136) argument pairs classified as “Against”.

Table 1. Excerpt of dataset sentence pairs, labeled according to the relation from Arg2 with Arg1

Arg1	Arg2	Relation
I think that using animals for different kind of experience is the only way to test the accuracy of the method or drugs. I cannot see any difference between using animals for this kind of purpose and eating their meat.	And I think there are alternative solutions for the medical testing. For example, for some research, a volunteer may be invited.	Against
I don't think the animal testing should be banned, but researchers should reduce the pain to the animal.	Maybe we should create instructions to follow when we make tests on animals. Researchers should use pain killer just not to torture animals.	InFavour
I think that using animals for different kind of experience is the only way to test the accuracy of the method or drugs. I cannot see any difference between using animals for this kind of purpose and eating their meat.	Animals are not able to express the result of the medical treatment but humans can.	Against

To build our classifier model we represented each sentence pair with a Bag-of-Words (BOW) and we added some calculated features aiming to improve the results of the classifier as the authors done in [17]. The calculated features consisted in Sentence Similarity [23], Edit Distance [24], and Sentiment Score [25] measures. Sentence Similarity and Edit Distance measures were calculated for each pair of argumentative sentences, while Sentiment Score was calculated individually to each argumentative sentence.

The experiments on building the classifier were carried out using two algorithms: Support Vector Machines (SVM) and Random Forest (RF). We also used a k-fold cross-validation procedure to evaluate our models due to the small size of dataset. The k parameter of k-fold was set to 10 and we obtained the results presented in **Table 2**.

Table 2. 10-fold CV average results on training dataset

	Accuracy(%)	Precision(%)	Recall(%)	F_1(%)
RF	56,21	60,04	55,38	53,96
SVM	66,07	65,69	69,22	64,22

4 Discussion

The results obtained are not yet satisfactory enough so that we can consider that these models can already be applied in the GDSS prototype that we have been developing, however as we can see in Table 2, the model generated by the SVM algorithm presents higher values in all measures model evaluation. As next steps we intend, first to improve our classifier performing feature selection analysis on the dataset in order to reduce the BOW to the main features (words) that can have more influence in the model definition and second to test our classifier with other debate datasets.

We also intend to continue working on the creation of the classification models, based on datasets generated with our argumentation framework, in this case we will have more features added to our dataset, such as: data on the decision makers' behavior style in the decision-making process, data on the decision-maker preferences on each of the alternatives and criteria of the problem, among others, which we believe could be usefull to construct more precise classifiers.

5 Conclusions

Supporting and representing decision-makers in group decision making processes is a complex task, but when we consider that the decision-makers may not be in the same place at the same time, the task is even more complex. The utilization of artificial intelligence mechanisms in the conception of GDSS's is boosting the improvement and acceptance of these systems by the decision-makers.

In this work we applied ML supervised classification algorithms to arguments extracted from online debates with the goal of creating an argumentative sentence classifier to perform automatic classification of argumentative sentences exchanged in GDSS's by decision-makers.

Although the results obtained do not have a high accuracy, we believe that it is possible to automatically classify the argumentative sentences exchanged by decision-makers during a group decision-making process with greater precision, when we obtain more features that help to relate the argumentative sentences. Those features can be extracted or calculated from the original sentences and others will be generated by our dynamic argumentation framework. Despite that, we still plan to study deeply the feature selection techniques in the natural language process and run experiences to compare the results.

Acknowledgments

This work was supported by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within the Projects UIDB/00319/2020, UIDB/00760/2020, GrouPlanner Project (POCI-01-0145-FEDER-29178) and by the Luís Conceição Ph.D. Grant with the reference SFRH/BD/137150/2018.

References

1. Bell, D.E.: Disappointment in decision making under uncertainty. Operations research 33, 1-27 (1985)
2. Huber, G.P.: A theory of the effects of advanced information technologies on organizational design, intelligence, and decision making. Academy of management review 15, 47-71 (1990)
3. Luthans, F., Luthans, B.C., Luthans, K.W.: Organizational Behavior: An evidencebased approach. IAP (2015)
4. Huber, G.P.: Issues in the Design of Group Decision Support Systems. MIS Quarterly: Management Information Systems 8, 195-204 (1984)
5. DeSanctis, G., Gallupe, B.: Group decision support systems: a new frontier. SIGMIS Database 16, 3-10 (1985)
6. Marreiros, G., Santos, R., Ramos, C., Neves, J.: Context-Aware Emotion-Based Model for Group Decision Making. Intelligent Systems, IEEE 25, 31-39 (2010)
7. Conceição, L., Martinho, D., Andrade, R., Carneiro, J., Martins, C., Marreiros, G., Novais, P.: A web-based group decision support system for multicriteria problems. Concurrency and Computation: Practice and Experience e5298 (2019)
8. Carneiro, J., Andrade, R., Alves, P., Conceição, L., Novais, P., Marreiros, G.: A Consensus-based Group Decision Support System using a Multi-Agent MicroServices Approach. In: International Conference on Autonomous Agents and Multi-Agent Systems 2020. International Foundation for Autonomous Agents and Multiagent Systems, (2020)
9. Grudin, J.: Group Dynamics and Ubiquitous Computing. Communications of the ACM 45, 74-78 (2002)
10. Carneiro, J., Martinho, D., Marreiros, G., Jimenez, A., Novais, P.: Dynamic argumentation in UbiGDSS. Knowledge and Information Systems 55, 633-669 (2018)
11. Carneiro, J., Martinho, D., Marreiros, G., Novais, P.: Arguing with Behavior Influence: A Model for Web-based Group Decision Support Systems. International Journal of Information Technology & Decision Making (2018)
12. Carneiro, J., Saraiva, P., Martinho, D., Marreiros, G., Novais, P.: Representing decision-makers using styles of behavior: An approach designed for group decision support systems. Cognitive Systems Research 47, 109-132 (2018)

13. Carneiro, J., Saraiva, P., Conceição, L., Santos, R., Marreiros, G., Novais, P.: Predicting satisfaction: Perceived decision quality by decision-makers in Web-based group decision support systems. Neurocomputing (2019)
14. Lawrence, J., Reed, C.: Argument Mining: A Survey. Computational Linguistics 45, 765-818 (2020)
15. Benlamine, S., Chaouachi, M., Villata, S., Cabrio, E., Frasson, C., Gandon, F.: Emotions in argumentation: an empirical evaluation. In: Twenty-Fourth International Joint Conference on Artificial Intelligence. (2015)
16. Rosenfeld, A., Kraus, S.: Providing arguments in discussions on the basis of the prediction of human argumentative behavior. ACM Transactions on Interactive Intelligent Systems (TiiS) 6, 30 (2016)
17. Carstens, L., Toni, F.: Towards relation based argumentation mining. In: Proceedings of the 2nd Workshop on Argumentation Mining, pp. 29-34. (2015)
18. Cocarascu, O., Toni, F.: Identifying attack and support argumentative relations using deep learning. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1374-1379. (2017)
19. Rosenfeld, A., Kraus, S.: Strategical argumentative agent for human persuasion. In: Proceedings of the Twenty-second European Conference on Artificial Intelligence, pp. 320-328. IOS Press, (2016)
20. Swanson, R., Ecker, B., Walker, M.: Argument mining: Extracting arguments from online dialogue. In: Proceedings of the 16th annual meeting of the special interest group on discourse and dialogue, pp. 217-226. (2015)
21. Mayer, T., Cabrio, E., Lippi, M., Torroni, P., Villata, S.: Argument Mining on Clinical Trials. In: COMMA, pp. 137-148. (2018)
22. Lippi, M., Torroni, P.: MARGOT: A web server for argumentation mining. Expert Systems with Applications 65, 292-303 (2016)
23. Miller, G.A.: WordNet: a lexical database for English. Communications of the ACM 38, 39-41 (1995)
24. Navarro, G.: A guided tour to approximate string matching. ACM computing surveys (CSUR) 33, 31-88 (2001)
25. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: LREC, pp. 417-422. Citeseer, (2006)

2.3 Aspect Based Sentiment Analysis Annotation Methodology for Group Decision Making Problems: An Insight on the Baseball Domain

Title	Aspect-Based Sentiment Analysis Annotation Methodology for Group Decision-Making Problems: An Insight on the Baseball Domain
Authors	Tiago Cardoso, Vasco Rodrigues, Luís Conceição, João Carneiro, Goreti Marreiros and Paulo Novais
Publication Type	Conference Proceedings
Publication Name	Lecture Notes in Networks and Systems (LNNS)
Publisher	Springer, Cham
Volume	469
Pages	25-36
Year	2022
Month	May
Online ISBN	978-3-031-04819-7
Print ISBN	978-3-031-04818-0
URL	https://link.springer.com/chapter/10.1007/978-3-031-04819-7_3
State	Published

Aspect Based Sentiment Analysis Annotation Methodology for Group Decision Making Problems: an insight on the Baseball Domain

Tiago Cardoso[1], Vasco Rodrigues[1], Luís Conceição[1,2], João Carneiro[1], Goreti Marreiros[1] and Paulo Novais[2]

[1] GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Institute of Engineering, Polytechnic of Porto, 4200-072 Porto, Portugal
[2]ALGORITMI Centre, University of Minho, Guimarães 4800-058, Portugal

Abstract. Decision making is an important part of our lives, especially in the context of an organization where decisions affect their business, and in this modern era, it is increasingly important to make the best decisions and increasingly difficult to get people together to make said decisions. Because of this, the importance of Group Decision Support Systems keeps growing, especially those that are web-based since they allow a connection between people in different corners of the world. However, there isn't much in terms of systems that can take online text-based discussions and use them to help a group of people reach a decision. This works addresses one of the aspects of this issue, that being the lack of annotated datasets that can provide a source of information to help in the creation of said systems. For this purpose, this work presents a methodology to be applied to unstructured text-based discussions found on the social web, to extract from them important information and organize it. In addition, a practical case study of this methodology is described, using Baseball domain discussions from Reddit as this case's unstructured data. We concluded that the created methodology allows the structuring of different aspects of a given social web discussion, especially in Reddit, and could be applied to discussions found on several existing domains.

Keywords: Group Decision Making, Unstructured Data, Text Annotation, Reddit, Baseball

1 Introduction

Decision making is an important part of day-to-day life, be it at an individual level or group level. In the case of group focused decisions, this process is referred to as group decision making (GDM) and it is seen as a complex thing mostly due to the personalities and interactions between the different decision makers [1].

2

Given how most of our day-to-day lives are organized, we are faced with several of these GDM situations, be it in a professional or social setting, it is believed that making decisions as a group tends to provide better outcomes, because a group can cooperate to solve a given problem, reducing the chances of mistakes being made. Still, for a good decision to be reached, the conditions need to be right to help maximize the group's chances of reaching a decision [2]. Specifically, it isn't easy to physically bring a group of people together to make a decision, this can be offset through the usage of Web-based Group Decision Support Systems (web-based GDSS) that while seen as important tools, haven't been widely accepted or used, mostly because of a resistance to change from organizations and a fear of losing the value obtained from physical meetings [1], [3].

Nevertheless, in the case of the text that comes from chatrooms or messages, it may be used to identify and analyze the intent behind it and how it may contribute to a decision-making process. Performing a study on other works related to Sentiment Analysis and decision-making, there are cases where text has been used to extract information [4], [5] yet for the most part these use data from websites or apps that allow users to leave quick text and numerical reviews. Meaning that there is a need for a system, that could process less structured and longer text usually exchanged during discussions and turn it into meaningful information. GDSS systems can be implemented in different ways, they can also be different in how they handle text and their learning style, with certain styles needing properly annotated datasets [6]. Considering this, a careful search for datasets that could be applied to a decision-making context was made, focusing on places like Kaggle[1] and NLP Index[2], that concluded that most of those available for the most part had quick reviews that wouldn't be enough in a GDM context, that needs discussion and argumentation about a topic [7].

After extensive searching and discussions, we agreed that a good source for argumentative text would be Reddit[3] since this platform enables the creation of discussions about any domain and the possibility of argumentation between its users. The downside being that there is no structured data that a system based on supervised learning could handle and here comes the problem that this paper intends to tackle [8]. There is a lack of datasets focused on the annotation of discussions, especially when most focus on quick text reviews, this lack of datasets reveals at the very least, a need for a comprehensive methodology to create this type of datasets.

Further study showed that there was an annotation structure that could provide a base for this methodology. This study was performed on Task 5 of the 2016 edition of the International Workshop on Semantic Evaluation (SemEval-2016[4]), that focused on Aspect Based Sentiment Analysis. More specifically, it focused on identifying opinions expressed in customer reviews about certain domains, so it served as a base that we could improve on to fit our needs. With the creation of this annotation meth-

[1] kaggle.com
[2] index.quantumstat.com
[3] reddit.com
[4] alt.qcri.org/semeval2016/

odology and its use to create datasets based on Reddit discussions, we intend to enable the creation of intelligent models that can be trained to identify important pieces of information that are relevant to the decision-making process and facilitate reaching an agreement.

The remainder of this work is organized into three other sections. Section 2 addresses methodology itself, with a focus on its structure, different Features and how it can be applied to different domains. In Section 3 we address the initial collection of data from Reddit and how it was treated so that the methodology could be applied to it, along with the specification of the relevant Features and the other important steps to create the dataset that is the focus of this work. Finally, Section 4 presents the conclusions of this work and some indications for future work.

2 Dataset Annotation Methodology

As a basis for the methodology, it was utilized the guidelines presented in SemEval-2016, more specifically the annotation guidelines of Task 5 [9] of this workshop, which establish the concept of Entity Types, Attribute Labels, Opinion Target Expression and Opinion Polarity.

While these 4 concepts were already used to characterize a sentence, they were not enough for the GDM context, so to better annotate the opinions and alternatives present in a sentence, other Features were formulated, like **Aspect**. This Feature was created to indicate if a certain Entity is indicated explicitly or not in a certain opinion. The values this annotation can take are Explicit or Implicit, and since a sentence can have more than one Entity Type., **Criteria and Criteria List**. The Criteria indicates if an opinion contains a Criterion to help in the decision-making problem, the value can be True or False, while Criteria List is the list of Criteria present in the text. Additionally, there are cases where there is not a criterion and the Criteria List doesn't have values, mostly the case of general opinions like "X is the best" where the person with that opinion does not state what makes X the best at something., and **Alternative and Alternative List**. Alternative indicates if the opinion contains an alternative to the decision-making problem, the value can be True or False. The Alternative List is a list of values that contains the identifier of the alternative, that being the name of the player(s) considered in the sentence. For example, "X is the greatest", the OTE is X while the Alternative List would have a normalized way of the OTE, be it a full name or some other established way of identifying X. with the intention of keeping track of the alternatives being discussed, which cannot be fully represented via the Opinion Target Expression since it would insufficient for a GDM problem. There is also a Feature, **Type**. The Type annotation indicates if a certain phrase depends or not on another one to make sense. A sentence can be classified as Main (M), if it establishes a new subject/Entity or is capable of existing without any sentence preceding it, if this isn't the case a sentence is classified as Dependent (D). This Feature can only have one value., used to indicate if a sentence is dependent on a previous sentence or not. Considering a finer analysis of a sentence, we should take into consideration that the same sentence can contain different opinions meaning that one sentence can have a single value, or a list of values associated with certain Features to fully characterize such sentence [10].

Having established the different Features, then guidelines were created to ease their usage. This being done with the creation of different questions to "ask the text" as it was being analyzed, to help, for example, identify the subject of the sentence, the Entity Types or relation between Attribute Label and Entity Type. Along with these questions, other rules were put in place to better understand what to do when considering specific sentence or grammatical structures.

These different aspects of the methodology will be explored in further detail in this section.

2.1 Features

- **Aspect**. This Feature was created to indicate if a certain Entity is indicated explicitly or not in a certain opinion. The values this annotation can take are Explicit or Implicit, and since a sentence can have more than one Entity Type.
- **Entity Types**. An Entity Type, or just Entity, is used to represent concrete characteristics of the topic being discussed [9], so that if the conversation focuses on restaurants, an Entity can be the restaurant or its service, indicating that it doesn't need to be a physical or palpable concept. The process of defining Entities always depends on the domain that is going to be annotated and requires some discussion to reach an agreement between everyone in terms of granularity and what should be considered.
- **Attribute Labels**. An Attribute Label, or simply Attribute, is used to characterize an Entity and because of that, it is not as clear or palpable concept as the latter [9], and it can be general in what characterizes or be more specific. In terms of establishing Attributes, the Entities should be clear and finalized first and the Attributes should make sense in conjugation with the Entities and the discussion that is taking place. The Entity-Attribute relationship can be characterized using a combination table that is a clear and comprehensive representation of the domain of the discussion in question.
- **Opinion Target Expression**. The Opinion Target Expression (OTE) is a clear reference to the Entity present in an opinion [9]. It can be a single value or a list of values, with these values being either the linguistic expression used to refer to the reviewed Entity, for example, a player's name (complete or not) or brand of a product or in the case of the Aspect of the sentence being Implicit, the value is NULL.
- **Criteria and Criteria List**. The Criteria indicates if an opinion contains a Criterion to help in the decision-making problem, the value can be True or False, while Criteria List is the list of Criteria present in the text. Additionally, there are cases where there is not a criterion and the Criteria List doesn't have values, mostly the case of general opinions like "X is the best" where the person with that opinion does not state what makes X the best at something.
- **Alternative and Alternative List**. Alternative indicates if the opinion contains an alternative to the decision-making problem, the value can be True or False. The Al-

ternative List is a list of values that contains the identifier of the alternative, that being the name of the player(s) considered in the sentence. For example, "X is the greatest", the OTE is X while the Alternative List would have a normalized way of the OTE, be it a full name or some other established way of identifying X.

- **Opinion Polarity**. Opinion Polarity, or simply Polarity, represents the polarity of an opinion towards an Entity-Attribute pair in a phrase [9] and it can be Positive, Negative or Neutral. While Positive and Negative polarities are easy to understand and apply, Neutral polarity is rarely used, being in most cases used when an opinion can't be seen as fully Positive or Negative. For example, in "X is the best, but not the greatest", there is no specific stance on what the opinion is, so a Neutral polarity can be used.
- **Type**. The Type annotation indicates if a certain phrase depends or not on another one to make sense. A sentence can be classified as Main (M), if it establishes a new subject/Entity or is capable of existing without any sentence preceding it, if this isn't the case a sentence is classified as Dependent (D). This Feature can only have one value.

2.2 Comparison and Validation

In this step, annotations are compared to each other to detect differences, when differences were found a comparison step was needed. Each side should explain the line of thought use to reach that annotation, in order to understand what the cause for the difference is, since typing mistakes in the annotating process can cause divergences, but these are easily fixed.

If after the explanation of each side, a consensus isn't reached, the phrase should be marked for validation. In the validation step, someone aside from the annotators should hear both sides and with external reasoning, help decide on which makes more sense or come up with a new annotation all together.

People with good knowledge of the language used in the sentence are among the best to mediate this step since most of the differences come from different understandings of the same sentence [11]. This was exemplified in [12] where they used a native speaker linguistics student to annotate and then it was revised and corrected by a native and a non-native French speaker.

2.3 Guidelines

After establishing the Features to be annotated, some questions were defined to be able to identify the Entities and Attributes of a phrase and if an Entity is mentioned explicitly or not. Those questions are the following:

- To help determine the Entity Feature: "What is the Entity?"
- To help determine the Aspect Feature: "Is the Entity mentioned in the sentence or it is implied?"
- To help determine the Attribute Feature: "What is the mentioned Attribute, in relation to the Entity?"

- To help determine the Polarity Feature: “What is the sense of the sentence?”
- To help determine the OTE Feature: “What is the subject of the sentence?”

Few other guidelines:

- Neutral opinions normally are discarded since they normally do not help solve the decision-making problems.
- When in a certain opinion there is a direct comparison between players or their capabilities, there should be annotated all players with opposite polarities. For example, in “X is better than Y and Z” X should have a positive polarity while Y and Z should have a negative polarity.
- Whenever the word “but” is utilized, normally it means a change of polarity in the annotations. So, in “X is very good at running but he is not the greatest”, the polarity of the first part is positive while in the second one it becomes negative.
- The term “barely” can be used to describe an Entity in multiple ways depending on its use, needing careful deliberation before reaching a conclusion about an annotation. For example, in “X was so far outpacing the competition it was barely fair.” denotated something positive about the player X. In the case of “The only sprinter with a lower time is X, which is barely lower than Y.” the term is used to show how insignificant the difference between times is, making that criterion unviable.
- One phrase can have both Explicit and Implicit Aspect because of cases where an Entity’s name is referred, and it is followed by different references to it in a way that the phrase can be divided without the name being in it. Examples of this can be the use of subject pronouns like “X did not play for long time but when he did, it was amazing” or the use of possessive pronouns for example, “X did amazing times in the 100m but his times on 200m were not as good”.
- Sentences with Explicit Aspect can be Dependent if they need the context mentioned previously to be coherent.
- A Dependent Sentence should always have a Main Sentence associated to it. If the Main Sentence is not relevant for the context of the discussion, subsequent Dependent Sentences should be removed even if they are relevant.

3 Case Study

Next, we focus on the application of the methodology, the definition of the relevant Features, collection, and treatment of data from Reddit, detail some of the annotation process and the creation of the final dataset.

3.1 Intermediary Dataset Creation

Data Extraction. For the purposes of this work, it was decided that one of the better sources of argumentative, yet non-formal text would be Reddit, since it is a website that hosts many themed communities called subreddits, where different discussions take place. In this context, a focus was placed in finding discussions based around determining the "best" in a given domain (e.g., Baseball, MotoGP, or Movies), since these usually bring out the most argumentative side of a community.

A search was performed over different subreddits, with the objective of finding a preferably well worded discussion with several replies, that were justified in the opinions given. After a search through these different subreddits, it was finally decided to settle on "r/baseball"[5], a subreddit dedicated to the sport of Baseball, from where two discussions were chosen, that being "Who do you think is the greatest baseball player of all-time?"[6] that was designated as "Discussion A" and "Question on the greatest baseball player ever."[7] that was designated as "Discussion B".

Having identified the information that would be used for the creation of our dataset, the discussions were extracted from Reddit on 2021/07/07, by downloading them under the format of a json files.

Data Handling. For both discussions, their respective json files had to be treated and turned into a format that could be considered "useful". To achieve this, both json files were put through a process that isolated the relevant pieces of information and organized them into preliminary datasets.

The first preliminary dataset focused on for every post identifying the Discussion it came from (*dataset*), identifying its author (*userid*), itself (*messageid*), identifying its "parent" post if it exists (*originalmessageid*), associated positive (*messageups*), negative (*messagedowns*) and overall (*messagescore*) scores, and its content (*messagetext*).

The second preliminary dataset is very similar to the previous, however for the annotation process, the content of a post was split into the sentences that compose it. This led to the addition of an identifier of a sentence that is made up of a post's id and a number indicating the location of the sentence (*sentenceid*) and the content of the sentence (*sentencetext*).

After the creation of the second preliminary dataset the content of the posts had to be treated, to make them easier to handle. For this, each sentence was preprocessed with the usual treatment for Reddit data for example, the normalization of HTML characters like "&amp" and "&#x200b" and removal of URLs [13].

3.2 Features

Entity Types. After a study of the Baseball domain, we decided to create the following Entities:

- **Player**: Used to represent opinions relative to the player itself. Related to a player's talent, impact, achievements, etc.

[5] reddit.com/r/baseball
[6] reddit.com/r/baseball/comments/2r5imp/who_do_you_think_is_the_greatest_baseball_player
[7] reddit.com/r/baseball/comments/av3gcy/question_on_the_greatest_baseball_player_ever

- **Stats**: Used to represent opinions relative to the player's stats like Homeruns, Career totals, etc.
- **Pitching**: Used to represent opinions relative to the player's pitch
- **Hitting**: Used to represent opinions relative to the player's hitting. It takes into consideration things like a player's swing, bat speed, hitting power, etc.
- **Physical Condition**: Used to represent opinions relative to the player's physical condition. Related with a player's health problems, physical qualities, usage of performance-enhancing drugs (PED), etc.
- **Offensive**: Used to represent opinions relative to the player's offensive skills like base running, speed, etc.
- **Defensive**: Used to represent opinions relative to the player's defensive skills like defense, fielding, etc.

Attribute Labels. Taking into consideration the Entities described earlier, it was decided to create the following Attributes:

- **General**: Used for situations where the opinion is generalized and something non-specific about the Entity
- **Ability**: Used for opinions related to the capability of a player to do an Entity, like a certain player being able or not to do Pitching
- **Performance**: For situations where something specific of a certain Entity is referred in the opinion, like bat speed in hitting or the quality in which a certain Entity is done or how the player performs, this attribute is utilized
- **Miscellaneous**: When the other Attributes cannot be utilized in a certain opinion, there is this attribute for those situations where no other Attributes can fit that opinion.

Entity-Attribute Combination Table. In the case of the Baseball domain, we considered the Entity-Attribute pairs shown in **Table 1**, where it can be seen how some Entities do not pair with certain Attributes, since such pairings do not make sense in the domain:

Table 1. Entity-Attribute Combinations

	General	Ability	Performance	Miscellaneous
Player	✓	✓	✓	✓
Stats	✓	×	×	✓
Pitching	✓	✓	✓	✓
Hitting	✓	✓	✓	✓
Physical Condition	✓	×	×	✓
Offensive	✓	✓	✓	✓
Defensive	✓	✓	✓	✓

Criteria and Criteria List. In the Baseball domain, with the Entities described earlier, we decided to maintain a consistency between the Entities and the Criteria, so that most Criteria would be the same as their Entity counterpart, except for the Player Entity since it is not a valid term for a criterion since it is too subjective. With that in mind, in the case of the Player Entity, we defined the following traits: Impact, Game Role, Era, Dominance, Career length, Achievements and Talent.

Alternative and Alternative List. Considering how both Discussion A and B are focused on the topic of the best Baseball Player, no Alternative List was created, instead this list was built as the annotation process went along based on what was said in the text. For example, in the sentence "Ruth is the best player of all time" the OTE is "Ruth", but the corresponding alternative is his full name, that being "Babe Ruth". Another consideration was in the case of nicknames where the corresponding alternative was the player's full name.

3.3 Specific Situations

During the annotation process there were cases when it was not easy to distinguish when to annotate an Attribute as General or Performance, even when these have distinct definitions.

These cases usually occurred when terms like "well" or "good" were used. The presence of these terms brought about these situations since at first glance they give the idea of the sentence discussing Performance, but with a better analysis of the context in which these terms are used, the Attribute General could instead be applied. There were also cases where the opposite also happened.

Another situation came from the fact that in both Reddit discussions there was a lot of weight placed on when a player was active (their era), or the state of the game in certain era, or era is mentioned through the "time machine scenario". In the "time machine scenario", a player from the past is brought to the present era to discuss and speculate on how they would react and adapt to the changes in the game.

While the uses of an era as an argument can be seen as different criteria to judge a player's merit, in the context of this work both usages were considered as part of the Era criterion. Of course, this approach lowers the granularity of the annotations, since different criteria could be used, however in the context of this work such granularity was seen as outside of the scope that we were trying to achieve.

3.4 Comparison and Validation

With the candidate annotations finalized by each annotator, in candidate annotations for Discussion A, from the 121 lines annotated, there was 69 differences and in Discussion B, from 397 lines, 193 were different. To solve these divergencies, we followed the Comparison and Validation mentioned earlier, where the 2 annotators reviewed the differences and if no consensus was reached in this process the issue was reviewed by the rest of the team.

The combination of these 2 processes was performed multiple times for both discussions to review and validate all the divergencies where both annotators couldn't reach a consensus on. By the end of the comparison and validation process two annotated datasets were created, one for Discussion A and another for Discussion B.

3.5 Dataset Finalization

The final process for this work was the creation of a final annotated dataset that contained the information of the annotated datasets representing Discussion A and B. For this merger to be made, the information of each dataset still needs further treatment.

The first consideration was to ignore the lines in the datasets that had been marked as irrelevant, since these were not considered as being relevant. As it was mentioned some Features can have more than one value, in these cases, these values need to be divided, so that the annotations of a given sentence can be turned into two or more lines in the dataset.

Another aspect that needed to be changed when creating the final dataset, was the column names or headers, since the final dataset will have 20 columns with different names from the 21 in the preliminary datasets. Column names were maintained except for Criterion which became *hasCriterion*, Alternative becoming *hasAlternative*, Criteria List and Alternative List that became *criterion* and *alternative* respectably. All column names being made to follow the camelCase codding standard[8].

3.6 Dataset Analysis

Analyzing the final dataset, we could see that majority of the final dataset data comes from Discussion B, totaling 77.5%, the Aspect mostly annotated was Explicit with 60.9%, with the participants of both Discussions establishing a new subject/Entity or engaging on the discussion with an opinion that is capable of existing without any sentence preceding it (Main Sentence - 60.9%). Diving into more feature analysis, in **Table 2**, we could see that the most used Entity to judge was Player with 59.8%, which leads to major usage of the General Attribute and most of the opinions annotated were of positive polarity (72.5%).

Table 2. Dataset Features Analysis

		Global Values	Percentages
Entity	Defensive:	12	2.5%
	Hitting:	34	7.0%
	Offensive:	19	3.9%
	Physical Condition:	32	6.6%
	Pitching:	29	5.9%

[8] techterms.com/definition/camelcase

	Player:	292	59.8%
	Stats:	70	14.3%
Attribute	General:	436	89.3%
	Ability:	7	1.5%
	Performance:	44	9.0%
	Miscellaneous:	1	0.2%
Polarity	Positive:	354	72.5%
	Negative:	131	26.9%
	Neutral:	3	0.6%

As mentioned earlier, to maintain a consistency between the Entities and the Criteria, most Criteria is the same as their Entity counterpart except the Player Entity since it is so subjective. From the 59.8% usage of the Player Entity, 40.2% does not use a Criterion to describe the Player, stating things like "He is the greatest" which doesn't have a Criterion to judge on, followed by 4.5% of the Impact Criterion and 3.3% of the Achievements Criterion.

An analysis of the OTE will not be performed since there are too many, considering the participants use the players' first name or last name or even nicknames. Instead, the analysis will be performed on the Alternative List, which is a normalized version of the OTE, in which we can understand that the participants talk more about Babe Ruth (30.9%), followed by Barry Bonds (16.4%) and Willie Mays (9.6%) making us conclude that the solution to this GDM problem lies in one of these Alternatives.

4 Conclusion

This work approached the creation of a dataset based on discussions acquired from a subreddit dedicated to Baseball, with these discussions focusing on the best Baseball player ever. To facilitate the annotation process of different Features, guidelines were created, based on the annotation guidelines of Task 5 of the SemEval-2016 workshop and discussions had among the team involved in this work.

Through this work and the established methodology, we intended to devise an approach to the problem of turning unstructured discussions into structured annotated data, which is needed for the GDM domain. So, with this work along with creating a new dataset, we also provide an adaptable methodology that can be adjusted to other domains to annotate similar discussions which can lead to resolution of the general shortage of structured datasets for the training not only present in GDM context but in any sort of Deep Learning applications that need structured datasets.

For future works, we intend on using this dataset for an initial phase of information extraction to better understand the different aspects of the dataset. Another point of interest for this dataset is its usage to train and test different Machine Learning models to be able to, for example, determine someone's preference for an alternative or criterion, and the creation of clusters, with focus on grouping the different participants in the discussions based on their preference for an alternative or criteria.

Acknowledgments. This work was supported by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) with GrouPlanner Project (POCI-01-0145-FEDER-29178) and within the R&D Units Project Scope: UIDB/00319/2020, UIDB/00760/2020, UIDP/00760/2020 and the Luís Conceição Ph.D. Grant with the reference SFRH/BD/137150/2018.

References

1. M. Tang and H. Liao, "From conventional group decision making to large-scale group decision making: What are the challenges and how to meet them in big data era? A state-of-the-art survey," Omega, vol. 100, p. 102141, Apr. 2021, doi: 10.1016/j.omega.2019.102141.
2. J. Carneiro, P. Alves, G. Marreiros, and P. Novais, "Group decision support systems for current times: Overcoming the challenges of dispersed group decision-making," Neurocomputing, vol. 423, pp. 735–746, Jan. 2021, doi: 10.1016/j.neucom.2020.04.100.
3. L. Conceição, D. Martinho, R. Andrade, J. Carneiro, C. Martins, G. Marreiros, P. Novais, 'A web-based group decision support system for multicriteria problems', Concurr. Comput. Pract. Exp., vol. 33, no. 2, 2021, doi: 10.1002/cpe.5298.
4. C. Zuheros, E. Martínez-Cámara, E. Herrera-Viedma, and F. Herrera, "Sentiment Analysis based Multi-Person Multi-criteria Decision Making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews," Inf. Fusion, vol. 68, pp. 22–36, Apr. 2021, doi: 10.1016/j.inffus.2020.10.019.
5. A. Nawaz, A. A. Awan, T. Ali, and M. R. R. Rana, "Product's behaviour recommendations using free text: an aspect based sentiment analysis approach," Clust. Comput., vol. 23, no. 2, pp. 1267–1279, Jun. 2020, doi: 10.1007/s10586-019-02995-1.
6. U. S. Shanthamallu, A. Spanias, C. Tepedelenlioglu, and M. Stanley, "A brief survey of machine learning methods and their sensor and IoT applications," in 2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA), Larnaca, Aug. 2017, pp. 1–8. doi: 10.1109/IISA.2017.8316459.
7. G. Marreiros, P. Novais, J. Machado, C. Ramos, and J. Neves, "A Formal Approach to Argumentation in Group Decision Scenarios," Inf. Manag. Syst., vol. 22, no. 2, pp. 165–198, 2006.
8. E. Cambria, B. Schuller, Y. Xia and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," in IEEE Intelligent Systems, vol. 28, no. 2, pp. 15-21, March-April 2013, doi: 10.1109/MIS.2013.30.
9. "absa2016_annotationguidelines.pdf." Accessed: Dec. 04, 2020. [Online]. Available: https://alt.qcri.org/semeval2016/task5/data/uploads/absa2016_annotationguidelines.pdf
10. P. Lambert, "Aspect-Level Cross-lingual Sentiment Classification with Constrained SMT," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Beijing, China, Jul. 2015, pp. 781–787. doi: 10.3115/v1/P15-2128.

11. D. Maynard and K. Bontcheva, ‘Challenges of evaluating sentiment analysis tools on social media’, Proc. 10th Int. Conf. Lang. Resour. Eval. Lr. 2016, pp. 1142–1148, May 2016, [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2016/summaries/188.html.
12. M. Apidianaki, X. Tannier, and C. Richart, ‘Datasets for aspect-based sentiment analysis in French’, in Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016, Jan. 2016, pp. 1122–1126, [Online]. Available: https://hal.archives-ouvertes.fr/hal-01838536.
13. Z. Jiang, S. I. Levitan, J. Zomick, and J. Hirschberg, “Detection of Mental Health from Reddit via Deep Contextualized Representations,” 2020, pp. 147–156, doi: 10.18653/v1/2020.louhi-1.16.

2.4 Supporting argumentation dialogues in Group Decision Support Systems: an approach based on dynamic clustering

Title	Supporting Argumentation Dialogues in Group Decision Support Systems: An Approach Based on Dynamic Clustering
Authors	Luís Conceição, Vasco Rodrigues, Jorge Meira, Goreti Marreiros, and Paulo Novais
Publication Type	Journal
Publication Name	Applied Sciences
Publisher	MDPI
Volume	12
Number	21
Year	2022
Month	December
Online ISSN	2076-3417
URL	https://onlinelibrary.wiley.com/doi/epdf/10.1002/cpe.5298
State	Published
SJR impact factor (2021)	0.51, Engineering, Miscellaneous (Q2)
JCR impact factor (2021)	2.838, Engineering, Multidisciplinary (Q2)

applied sciences

MDPI

Article

Supporting Argumentation Dialogues in Group Decision Support Systems: An Approach Based on Dynamic Clustering

Luís Conceição [1,2,*], Vasco Rodrigues [1], Jorge Meira [1], Goreti Marreiros [1] and Paulo Novais [2]

1 GECAD—Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, LASI—Intelligent Systems Associate Laboratory, Institute of Engineering, Polytechnic of Porto, 4200-072 Porto, Portugal
2 ALGORITMI Centre, LASI—Intelligent Systems Associate Laboratory, University of Minho, 4800-058 Guimarães, Portugal
* Correspondence: msc@isep.ipp.pt

Abstract: Group decision support systems (GDSSs) have been widely studied over the recent decades. The Web-based group decision support systems appeared to support the group decision-making process by creating the conditions for it to be effective, allowing the management and participation in the process to be carried out from any place and at any time. In GDSS, argumentation is ideal, since it makes it easier to use justifications and explanations in interactions between decision-makers so they can sustain their opinions. Aspect-based sentiment analysis (ABSA) intends to classify opinions at the aspect level and identify the elements of an opinion. Intelligent reports for GDSS provide decision makers with accurate information about each decision-making round. Applying ABSA techniques to group decision making context results in the automatic identification of alternatives and criteria, for instance. This automatic identification is essential to reduce the time decision makers take to step themselves up on group decision support systems and to offer them various insights and knowledge on the discussion they are participating in. In this work, we propose and implement a methodology that uses an unsupervised technique and clustering to group arguments on topics around a specific alternative, for example, or a discussion comparing two alternatives. We experimented with several combinations of word embedding, dimensionality reduction techniques, and different clustering algorithms to achieve the best approach. The best method consisted of applying the KMeans++ clustering technique, using SBERT as a word embedder with UMAP dimensionality reduction. These experiments achieved a silhouette score of 0.63 with eight clusters on the baseball dataset, which wielded good cluster results based on their manual review and word clouds. We obtained a silhouette score of 0.59 with 16 clusters on the car brand dataset, which we used as an approach validation dataset. With the results of this work, intelligent reports for GDSS become even more helpful, since they can dynamically organize the conversations taking place by grouping them on the arguments used.

Keywords: group decision making; dynamic clustering; natural language processing; argumentation

check for updates

Citation: Conceição, L.; Rodrigues, V.; Meira, J.; Marreiros, G.; Novais, P. Supporting Argumentation Dialogues in Group Decision Support Systems: An Approach Based on Dynamic Clustering. *Appl. Sci.* **2022**, *12*, 10893. https://doi.org/10.3390/app122110893

Academic Editors: Kuo-Ping Lin, Chien-Chih Wang, Chieh-Liang Wu and Liang Dong

Received: 2 October 2022
Accepted: 22 October 2022
Published: 27 October 2022

1. Introduction

Currently, most decisions made by higher-ups in organizations are made in groups [1]. Group decision making (GDM) is a process where a group of people, usually called decision makers, select one or more alternatives to solve a specific problem they are discussing. Typically, this is a procedure where the decision makers discuss their viewpoints and opinions to achieve a consensus. There are several advantages associated with group decision-making processes, such as improving the quality of the decision made or sharing the workload. Nevertheless, the right conditions need to be acquired to take advantage of this process, such as the possibility of interaction between decision makers, allowing them to exchange ideas and the ability to understand the reasoning behind different preferences [2–4].

Group decision support systems (GDSSs) have been widely studied over the recent decades to create these conditions, so that decision makers can decide effectively. Due to globalization, higher-ups have been dispersed throughout the globe with different time zones, which led to the creation of GDSSs that could solve these location and time issues, the web-based GDSS. The purpose of these systems is to support the group decision-making process by creating the necessary conditions for it to be effective and by allowing the process to be carried out from any place and at any time [5]. These systems provide a way to solve problems, one of the problems being multi-criteria problems.

Multi-criteria problems consist of a finite number of alternatives, which are known initially before solving the problem [6], and a limited set of criteria that enable comparing alternatives. These criteria should be accurate, coherent with the decision and its domain, feasible, independent of each other, and measurable [7]. When the definition of alternatives and criteria are completed, a multi-criteria decision-making problem is created, and decision makers can express their opinion, arguing on the available alternatives, valuable through the criteria set.

In GDSS, argumentation is ideal, since it makes it easier to use justifications and explanations in interactions between decision makers, allowing them to express their ideas clearly. In addition to that, argumentation can be used to influence their preferences and, subsequently, the outcome of the decision-making process, as well as aiding the creation of higher quality agreements and, at the same time, decreasing the number of unsuccessful negotiations [8]. It is essential to notice that these systems based on argumentation dialogues can generate vast amounts of information, since a group decision-making process usually spans several iterations (rounds), making it difficult for the decision makers to analyze and follow the decision-making process. In addition to that, these systems are not widely accepted in organizations due to several factors, such as the resistance to change from organizations and the fear of losing the value obtained from physical meetings, but mainly due to the lack of explanations on how the system is proposing such solutions [9,10].

To make GDSSs more appealing to organizations, machine learning is beginning to gradually be used in GDSS to enhance their capabilities, for example, through argument mining (AM). AM consists of the automatic identification and extraction of the structure of inference and reasoning expressed as arguments presented in the natural language [11]. AM is helping GDSS to become more attractive to organizations by automatically obtaining meaningful information from unstructured text, such as aspect terms, aspect categories, and polarity detection field [12]. AM can automatically extract data from the discussion's unstructured text natural language and then present those data to decision makers or, for instance, highlight the relevant messages. These improvements decrease the time participants must take to set up their preferences in a GDSS.

AM combines different fields of natural language processing, such as information extraction, knowledge representation, and discourse analysis [13]. In addition to those fields, sentiment analysis can also be performed in AM, be it a document, sentence, or aspect level with different outputs, binary (positive or negative), or multi-level [12]. Sentiment analysis performed at the aspect level is called aspect-based sentiment analysis (ABSA) and intends to classify opinions at the aspect level and identify the elements of an opinion [12].

To make GDSS more accessible to its users, the automatic identification of elements of an opinion is a step needed to reduce the time necessary for decision makers to set themselves up on these systems. For example, the automatic identification of alternatives and criteria on natural language text used in discussions should be a feature in future GDSSs to make them more acceptable to organizations. Some work has been done in the GDM context to achieve this solution. For example, machine learning classifiers can automatically classify the direction (relation) between two arguments [14] or create intelligent reports where an algorithm selects which information topics should be reported to decision makers. These features improve decision makers' perception of the problem they are deciding on through the ability to present accurate and relevant information [14]. All these advancements in AM applied to the GDM context will lead to a better understanding

Group decision support systems (GDSSs) have been widely studied over the recent decades to create these conditions, so that decision makers can decide effectively. Due to globalization, higher-ups have been dispersed throughout the globe with different time zones, which led to the creation of GDSSs that could solve these location and time issues, the web-based GDSS. The purpose of these systems is to support the group decision-making process by creating the necessary conditions for it to be effective and by allowing the process to be carried out from any place and at any time [5]. These systems provide a way to solve problems, one of the problems being multi-criteria problems.

Multi-criteria problems consist of a finite number of alternatives, which are known initially before solving the problem [6], and a limited set of criteria that enable comparing alternatives. These criteria should be accurate, coherent with the decision and its domain, feasible, independent of each other, and measurable [7]. When the definition of alternatives and criteria are completed, a multi-criteria decision-making problem is created, and decision makers can express their opinion, arguing on the available alternatives, valuable through the criteria set.

In GDSS, argumentation is ideal, since it makes it easier to use justifications and explanations in interactions between decision makers, allowing them to express their ideas clearly. In addition to that, argumentation can be used to influence their preferences and, subsequently, the outcome of the decision-making process, as well as aiding the creation of higher quality agreements and, at the same time, decreasing the number of unsuccessful negotiations [8]. It is essential to notice that these systems based on argumentation dialogues can generate vast amounts of information, since a group decision-making process usually spans several iterations (rounds), making it difficult for the decision makers to analyze and follow the decision-making process. In addition to that, these systems are not widely accepted in organizations due to several factors, such as the resistance to change from organizations and the fear of losing the value obtained from physical meetings, but mainly due to the lack of explanations on how the system is proposing such solutions [9,10].

To make GDSSs more appealing to organizations, machine learning is beginning to gradually be used in GDSS to enhance their capabilities, for example, through argument mining (AM). AM consists of the automatic identification and extraction of the structure of inference and reasoning expressed as arguments presented in the natural language [11]. AM is helping GDSS to become more attractive to organizations by automatically obtaining meaningful information from unstructured text, such as aspect terms, aspect categories, and polarity detection field [12]. AM can automatically extract data from the discussion's unstructured text natural language and then present those data to decision makers or, for instance, highlight the relevant messages. These improvements decrease the time participants must take to set up their preferences in a GDSS.

AM combines different fields of natural language processing, such as information extraction, knowledge representation, and discourse analysis [13]. In addition to those fields, sentiment analysis can also be performed in AM, be it a document, sentence, or aspect level with different outputs, binary (positive or negative), or multi-level [12]. Sentiment analysis performed at the aspect level is called aspect-based sentiment analysis (ABSA) and intends to classify opinions at the aspect level and identify the elements of an opinion [12].

To make GDSS more accessible to its users, the automatic identification of elements of an opinion is a step needed to reduce the time necessary for decision makers to set themselves up on these systems. For example, the automatic identification of alternatives and criteria on natural language text used in discussions should be a feature in future GDSSs to make them more acceptable to organizations. Some work has been done in the GDM context to achieve this solution. For example, machine learning classifiers can automatically classify the direction (relation) between two arguments [14] or create intelligent reports where an algorithm selects which information topics should be reported to decision makers. These features improve decision makers' perception of the problem they are deciding on through the ability to present accurate and relevant information [14]. All these advancements in AM applied to the GDM context will lead to a better understanding of the conversations. They will result in a more intelligent way to organize and display the conversation to the decision maker, enhancing the capabilities of the GDSSs.

This work intends to apply unsupervised techniques, most concretely clustering, to group arguments to organize the discussion dynamically. This work's output will enhance the capabilities of the GDSS by offering an ML-based dynamic organization of ideas that could outperform a standard filter, since these standard filters need annotations to be viable. Unsupervised techniques ditch the need for annotated data. They can detect comparisons being made and group them, as well as detecting micro-discussions about a small group of alternatives and group them together.

The rest of the paper is organized in the following order: Section 2 presents some related works on clustering in natural language processing, Section 3 presents the methodology to solve the problem addressed in this article, Section 4 presents the obtained experiments results, and Section 5 presents a discussion about the obtained results. In the last section, some conclusions are presented, alongside suggestions for work to be done afterward.

2. Related Work

This section intends to provide insight into what has been done in terms of clustering with NLP data. Since the context of this work is particular, approaches from multiple fields will be explored, and a critical analysis of each one will be performed to understand if it could be applied to the GDM context. Many approaches were made before clustering on natural language text to achieve multiple objectives.

Kim et al. [15] applied clustering with NLP in the biology field by extracting data from two diverse sources, microarray gene expression data and gene co-occurrences in the scientific literature from bioRxiv using NLP. After normalizing the microarray data and applying dimensionality reduction with principal component analysis (PCA), they grouped this data into clusters using the K-means technique. The resulting clusters were compared to the extracted gene co-occurrences pairs in the NLP data to evaluate the results of the steps taken. The evaluation was done using entropy analysis on the combined data, comparing it to the maximum entropy from the sole clusters. Their results approve the usage of NLP in this field to extract gene co-occurrences from the literature in which the use of clustering helped confirm this claim. Although the approach shows an excellent combination of both areas, it is not what it is intended with this project, since the goal is to cluster unstructured text and not structured data, that was, in their case, the microarray gene expression data.

Sarkar et al. [16] applied clustering with NLP as an intermediary step in creating a model to predict occupational accident risk. After extracting the data from an integrated steel plant's safety management system database, pre-processing is done where duplicates, missing data, and inconsistent data are removed. The authors used EM-based text clustering to build clusters with categorical attributes while using the silhouette coefficient to determine the optimal number of clusters. These data are then fed to a deep neural network (DNN) model, with a structure comprised of a stacked autoencoder (SAE) with an autoencoder (AE) and a SoftMax classifier. The AE is a feed-forward artificial neural network (ANN) comprising one input layer, one hidden layer, and one output layer. Usually, it is trained to copy its input to its output so that the errors become minimum. Therefore, the dimension of the input must be the same as that of the output. Support vector machine and random forest was used to compare this approach. For DNN, the grid search technique was used to find the best hyperparameters. This approach shows one usage of clustering to categorize unstructured data, finding hidden connections between them.

Hema and David [17] applied clustering with NLP in the medical field as an intermediary step in creating a model to predict diseases based on symptoms. The data are collected using medical forums about various stomach disease symptoms, and an OWL file is created. After data preprocessing, stopwords, stemming words, special characters, numbers, and white spaces were removed. Speech tagging is used to extract verbs, nouns, subjective words, adverbs, etc., from the dataset, so afterward, Fuzzy c means can cluster the data into groups of common symptoms. RDF is then utilized for taxonomic relations, object relations, and data, while OWL is used for attribute relations. These relations are then mapped for the genetic algorithm to predict the disease of a customer based on the symptoms. This approach uses many steps that could be utilized in this project. However, having an intra-sentence segmentation step in our project, the usage of fuzzy c-means becomes less needed. Most of the data will be treated so that it can only be part of one cluster, removing the need for soft clustering techniques.

Dragos and Schmeelk [18] applied clustering with NLP in education to obtain meaningful information from student surveys. They receive open-text survey answers from five cybersecurity courses and cluster the answers to each question on the survey. They first select the cluster number based on heterogeneity. This measure represents the sum of all squared distances between data points in a cluster and the centroids. After the number of clusters is decided for each question, they use TF-IDF as word embedding to cluster the data with k-means and obtain categories based on the top keywords made manually. With this, they aim to fill the gap in identifying valid interpretations of student feedback in the literature. This approach was applied to education and student surveys. However, it seems like it can be adapted into any other field. Both techniques used are not domain-specific, and the categorization done afterward was manual according to the top keywords, meeting the objectives of our work in terms of clustering.

Gupta and Tripathy [19] applied clustering with NLP by creating a methodology that could be used in any domain. They tested it in a zoo dataset. The method consists of implementing a form of clustering that takes a non-numeric dataset and clusters it with the help of the word embeddings provided by the GloVe dataset by generating the vector representation for each of the sentences in the dataset of those words. Then, a dimensionality reduction is performed on the data set using t-distributed stochastic neighbour embedding (t-SNE) to obtain the accurate number of dimensions for proper cluster formation. The data are then clustered using k-means++. The only issue with this technique is that it chooses the number of clusters based on minimum inertia and the least number of clusters in total. They surpassed this difficulty by using the elbow method to decide the number of clusters formed by the algorithm. This methodology sounds interesting on paper, and the possibility of using it in any domain allows it to be adapted to this work.

Huang et al. [20] applied clustering with NLP in StackOverflow discussions to mine comparable technologies and opinions. They utilize tags in each discussion, considering the collection of technologies that a person would like to compare. To learn the tags, they compared two of the most used methods, the continuous skip-gram model and the CBOW model, where the first model outperforms the latter by a marginal difference. With this better model, they compared the difference between the number of dimensions and concluded that eight hundred was the one to use with the best accuracy. To obtain categorical knowledge, they run the tags against TagWiki to get its definition and then extract the tag category with a POS tagger. To mine comparative opinions, they extracted comparative sentences between both by using three steps for each pair of comparable technologies in the knowledge base. They first preprocessed the discussion considering only answers with a positive score and removing the punctuation and sentences that ended with question marks because they wanted to extract facts and not doubts. Finally, they lowercase everything to make tokens consistent with the technologies. Secondly, they locate candidate sentences using a large thesaurus of morphological forms of software-specific terms to match with tag names. In the last step, they select comparative sentences and develop a set of sentence patterns considering POS tags to obtain them. They use Word Mover's Distance to measure the similarity between sentences, which is helpful for short text comparison. This approach uses word embeddings to get a dense vector representation of each keyword from POS tags for comparisons, such as comparative adjectives and nouns, excluding the technologies under comparison. They then compute the minimal distance between keywords between sentences and use those distances in a similarity score. If the similarity is superior to the threshold, they are considered similar. Finally, to cluster representative comparison aspects, they build a graph where each node is a sentence. They use TF-IDF to extract keywords from a comparative sentence in one community to represent the comparison aspect of this community, removing stop words and choosing the top three with the highest scores to represent the community. Each community is regarded as a document. This approach starts to be more in line with the objectives of our project, especially on comparative opinions mining, which is like our goal of reaching a consensus on the best alternative in a particular problem using criteria to describe the available alternatives.

Y. Liu et al. [21] applied clustering with NLP by collecting COVID-19-related data from Reddit in subreddits of North Carolina, utilizing data preprocessing techniques, such as POS tagging and stop word removal. After this step, GloVe and Word2Vec embedders were tested with the cosine similarity measure used to calculate the similarity between words. Topic modelling techniques and a BERT model were fine-tuned to find people's concerns and key points from the sentences typed on Reddit posts. With the results of the last step, K-Means were used to cluster the sentence vectors into three categories, concluding that reopening and spreading the virus were the most discussed topics during the time of the posts gathered. Some aspects of this approach were considered in our work, such as data preprocessing options and word embedders.

Reimers et al. [22] applied clustering with NLP by testing it in the context of open-domain argument search. To classify and cluster topic-dependent arguments, they measure the quality of contextualized word embeddings, ELMo and BERT. In terms of argument clustering, twenty-eight topics related to current issues about technology and society were picked. Since argument pairs addressing the same aspect should be assigned a high similarity score and arguments on various aspects a low score, they used a weak supervision approach to balance the selection of argument pairs regarding their similarity. After handling this issue, agglomerative hierarchical clustering with average linkage was used to cluster arguments. They also tested K-means and DBSCAN but agglomerative hierarchical clustering provided the best results in preliminary experiments.

Färber and Steyer [23] applied clustering with NLP on the argument search domain to identify arguments in natural language texts. To present aggregated arguments to users based on topic-aware argument clustering, they tried K-means and HDBSCAN, in addition to considering the argmax of the TF-IDF and LSA vectors to evaluate the results. Regarding word embeddings, TF-IDF, and BERT models, Bert-avg and Bert-cls were used as a pre-step for the clustering task. Another interesting remark is that they evaluated whether calculating TF-IDF within each topic separately is superior to computing the overall arguments in the document corpus. The dimensionality reduction technique, UMAP, was tested before clustering to verify its performance related to not using it in which HDBSCAN outperforms k-means on Bert-avg embeddings but using UMAP in combination with TF-IDF results in a slightly reduced performance. They found that Bert-avg embeddings result in marginally better scores than Bert-cls when using UMAP, concluding that this methodology can mine and search for arguments from an unstructured text on any given topic. Reimers et al. [22] and Färber and Steyer [23] contributed to the field of argument searching, which is similar to our work but not in the same context. Their approaches were used as an example for our project, using context-aware word embedding models (ELMo, BERT, and TF-IDF), the clustering techniques used, and their tested hyperparameters. The dimensionality reduction aspect brought by Färber and Steyer [23] is also interesting, as it helped obtain better results by reducing the number of features passed to the clustering techniques.

Dumani and Schenkel [24] applied clustering with NLP by creating a quality-aware ranking framework for arguments extracted from texts and represented in graphs. To achieve that, they used a (claim, premise) dataset based on debates taken on online portals in which they used SBERT instead of BERT, previously used on [25], to obtain the embeddings of the claims and premises. With these embeddings, agglomerative clustering using Euclidian distance metric and average linkage method was applied to achieve the clustering task. Since the dataset was sizeable (400 k) with many dimensions from the embedder (1024 dimensions), to reduce the time it would take to cluster it with the agglomerative technique, they clustered the dataset with K-means for K = 4. Then, they used agglomerative clustering on the results of K-means. This approach brought to attention some interesting points, such as the size of the dataset used and which measures could be taken to overcome that. Instead of dimensionality reduction, they used K-means as a pre-clustering step to reduce the computational time.

Daxenberger et al. [26] applied clustering with NLP to the argument mining field by creating an argument classification and clustering project for generalized search scenarios. For that, the technology mines and clusters arguments from various textual sources for a broad range of topics and in multiple languages were used, generalizing to many different textual sources, ranging from news to reviews. After fine-tuning a BERT base model, since it outperforms the pre-trained variant by a good margin, the embeddings obtained by this model are sent to the agglomerative hierarchical clustering with a stopping threshold, aggregating all arguments retrieved for a topic into the clusters of aspects. This project, in terms of argument clustering, seems promising. They used a fine-tuned BERT model for the word embeddings and utilized agglomerative hierarchical clustering to obtain arguments divided by aspects, such as the one presented in our project.

3. Methodology

This section addresses the utilized datasets, the processing pipeline, and the definition of the experiments tested.

3.1. Datasets

The two datasets used for this work were created using the methodology for annotating aspect-based sentiment analysis datasets [27]. The baseball dataset discusses which player is the best of all time. The Cars dataset discusses which car brand is the best and why. Both were extracted from Reddit.

The baseball dataset offers 488 rows of annotated data with 20 features, while the car brands dataset offers 388 rows with 16 features. The baseball dataset is recent, which is why it has more features than the 16 categorical features of car brands. These new features consist of a categorical feature to tell from each discussion the row it came from and message upvotes, downvotes, and message score numerical features.

As explained in [27], from the features created, the following features will be used to achieve the proposed objective:

- Sentence text—message typed by a user divided into the sentence level;
- Alternative—list of values that contains the identifier of the alternative;
- Criterion—list of criteria present in the text;
- Aspect—indicates if a specific entity is indicated explicitly or not in a particular opinion, taking explicit or implicit values;
- Polarity—polarity of an opinion towards an entity–attribute pair in a phrase. It can be positive, negative, or neutral;
- OTE—an apparent reference to the entity present in an opinion.

3.2. Methods

This section addresses the processing pipeline: preprocessing tools, word embedders, dimensionality reduction techniques, and clustering techniques.

3.2.1. Preprocessing Steps

Since text documents are unstructured by nature, to properly use them in NLP, suitable preprocessing is needed to transform and represent those text documents in a more structured way so they can be used later on [28]. In addition to that, it increases the quality of the final results when applied to classification problems, clustering, and other types of issues [29,30].

Online user-generated content, such as forums and social media discussions, is increasingly important, since it can provide essential knowledge to companies and organizations. However, this type of content has lots of noise, such as abbreviations, non-standard spelling, a specific lexicon of the platform, and no punctuation. These problems reduce the effectiveness of NLP tools, hence, why data preprocessing is needed [31].

Multiple steps can be taken in the preprocessing task, such as sentence segmentation, also known as sentence boundary detection and sentence boundary disambiguation, which consists of segmenting large paragraphs and documents into the fundamental unit of text processing, that is, a sentence [32]. Other operations, such as lowercasing text data [33,34] and stop word removal, are usually applied in preprocessing step. Stopwords are a type of word that does not have any linguistic value. Since they are considered low information, removing them allows for focusing on the essential terms of a text document [33]. This task requires a list of stopwords to remove specific words for each natural language, and this list is already compiled [34]. Stemming consists of reducing inflection in words to their base form, which can help deal with sparsity issues and standardizing the text document's vocabulary [33]. Lemmatization works similarly to stemming, in terms of reducing inflected words into their root form but varies in the fact that it tries to do it correctly without crude heuristics, making sure the word that resulted from the lemmatization (lemma) belongs to the language [33,34].

Furthermore, tokenization and normalization are used to break a text document into tokens, commonly words, for more accessible text manipulation [29]. It includes all sorts of lexical analysis steps, such as removing punctuation, number, accents, extra spacing, removing or converting emojis and emoticons, spelling correction, removal of URLs and HMTL characters, etc. [28,33,34]. In this work, a custom-designed intra-sentence segmentation tool was used. In addition to the standard sentence segmentation tool features, this tool works inside the sentence level. It detects comparisons, using the annotations to improve the results when assigning them back to the row. It then applies lowercasing of the text data, tokenization, removal of punctuation, and stopwords. Lemmatization of the tokens is the last preprocessing step taken.

3.2.2. Word Embedders

Word Embeddings transform text data into numerical representation, the so-called vectorization. According to Goldberg [35], word embedding, also known as distributed representations of words, is the term used to represent the technique where individual words are represented into real-value vectors. These vectors often have a dimension number in the tens or even thousands scale. Each word is mapped to one vector, representing a sentence in a list of these vectors. The mapping of a word to a vector can be done through dictionaries. This is better than using sparse word representations on the scale of thousands or even millions of dimensions [36].

TfidfVectorizer (scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer, accessed on 27 September 2022) is a method in the scikit-learn framework that enables the conversion of raw documents into a matrix of TF-IDF features. TF-IDF is the combination of the term frequency (TF) metric that represents the number of times a term occurs in a document versus the total number of terms in a document. In contrast, the inverse document frequency (IDF) represents the number of documents that contain the term [37]. The setting tested for this word embedder that works best for this work was **ngram_range = (1,1)**, where the first value indicates the minimum amount of grams to take into consideration and the second value the maximum, in which, by making them (1,1), we are solely utilizing unigrams.

Word2Vec (github.com/tmikolov/word2vec, accessed on 26 September 2022) is a popular word-embedding technique, developed by Tomas Mikolov [38]. It provides two methods to achieve this task, either by using a continuous bag-of-words (CBOW) [39] or skip-gram model (SG) [38]. The CBOW method takes the context of each word as the input and tries to predict the word corresponding to the context, while the SG method inverts the CBOW method, using the target word as the input and trying to predict the context. According to the author, CBOW is faster and has better representations for more frequent words, while SG works well with a small amount of data and represents rare words well [40]. In addition, the original GitHub implementation, the Word2Vec method, is in a commonly used framework called Gensim (radimrehurek.com/gensim/models/word2vec, accessed on 26 September 2022). We used the **word2vec-google-news-300** (huggingface.co/fse/word2vec-google-news-300, accessed on 27 September 2022) pre-trained vectors with an activated binary mode max length equaling 200.

Global vectors for word representation (GloVe) is the unsupervised learning algorithm for this task created by Stanford (nlp.stanford.edu/projects/glove/, accessed on 26 September 2022) [41]. It is available on GitHub (github.com/stanfordnlp/GloVe, accessed on 26 September 2022), where they supply pre-trained models for the task, depending on the requirements. Using the pre-trained models means the GloVe model becomes a static dictionary, since we obtain the word and its vector representation by downloading a pre-trained model [42]. In our work, we used **glove.6B.200d** (nlp.stanford.edu/projects/glove/, accessed on 26 September 2022) with 6 B tokens, 400 K vocab, uncased, and 200 dimensions in which we maintained that dimensions preset (max length equaling 200).

FastText is a library for efficiently learning word representation and sentence classification created by Meta Research (opensource.fb.com, accessed on 26 September 2022; github.com/facebookresearch, accessed on 26 September 2022) [43]. It is available on GitHub (github.com/facebookresearch/fastText, accessed on 26 September 2022), where they offer their state-of-the-art model for English word vectors and word vectors for 157 additional languages. It diverges from Word2Vec by using subword information on word similarity tasks to improve its results. We used **crawl-300d-2M** (fasttext.cc/docs/en/english-vectors, accessed on 26 September 2022), which consists of 2-million-word vectors trained on common crawl with 600 B tokens, and we used a max length equaling 200.

Bidirectional encoder representations from transformers (BERT) is a language representation model released by Google. It considers the context when creating word and sentence-embedding vectors, where the exact two words can have two different vectors, [44,45]. We utilized the **BERT-Base** pre-trained model with 12 layers, 768 hidden states, 12 heads, and 110 M parameters found on their GitHub page (github.com/google-research/bert, accessed on 26 September 2022).

Sentence-BERT (SBERT) (github.com/UKPLab/sentence-transformers, accessed on 26 September 2022) is a modification of the original pre-trained BERT network that uses Siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine similarity [46]. We utilized the **all-MiniLM-L6-v2** (sbert.net/docs/pretrained_models, accessed on 26 September 2022) pre-trained model with 6 layers and 384 hidden states totaling 1 billion training pairs.

Embeddings from language models (ELMo) is a state-of-the-art NLP framework developed by AllenNLP (allenai.org/allennlp/software/elmo, accessed on 26 September 2022). ELMo's representations differ from traditional ones because each token is assigned a representation that is a function of the entire input sentence. This way, word vectors are learned functions of the internal states of a deep bidirectional language model (biLM), which is pre-trained on a large text corpus [47]. We utilized this model's third version (**v3**) (tfhub.dev/google/elmo/3, accessed on 26 September 2022).

3.2.3. Dimensionality Reduction Techniques

In addition to word embedding, dimensionality reduction techniques are also advised to improve the accuracy of the clustering when handling data with a high number of features, making it advantageous in terms of computational efficiency [48,49].

Principal component analysis (PCA) was initially invented by Pearson [50] and later independently developed and named by Hotelling [51,52]. It is a statistical process that converts a group of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. All principal components are orthogonal to each other. Each one is chosen in a way that represents most of the available variance, with the first component having the maximum variance in a way that it selects a subset of variables from a more extensive set, based on which original variables have the highest correlation with the principal amount [53,54]. PCA can be named differently depending on the field of application, whereas in the ML field, it is called PCA and uses singular value decomposition (SVD) (scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA, accessed on 26 September 2022) [55].

The t-distributed stochastic neighbor embedding (t-SNE) was initially developed by Roweis and Hinton [56]. They created the concept of stochastic neighbor embedding, and later Van Der Maaten and Hinton [57] proposed the t-distributed variant. This variant is a nonlinear technique that converts similarities between data points to joint probabilities and tries to minimize the Kullback–Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data. The t-SNE has a cost function that is not convex, meaning that with different initializations, we can get other results [57]. This technique has implementations in multiple technologies, making it widely available (lvdmaaten.github.io/tsne, accessed on 26 September 2022). The t-SNE implementation in scikit-learn uses the Barnes–Hut approximation algorithm, which relies on quad-trees or octa-tree, which makes the maximum number of dimensions that can be used with t-SNE three.

Uniform manifold approximation and projection (UMAP) was developed by McInnes et al. [58] with a theoretical framework based on Riemannian geometry and algebraic topology. It is based on three assumptions, the data are uniformly distributed on a Riemannian manifold, the Riemannian metric is locally constant (or can be approximated as such), and the manifold is locally connected. This way, it is possible to model the manifold with a fuzzy topological structure. The embedding is found by searching for a low-dimensional data projection with the closest possible equivalent fuzzy topological structure (umap-learn.readthedocs.io/en/latest/, accessed on 26 September 2022) [58].

In addition to these three techniques, a hybrid approach applies PCA and t-SNE. PCA reduced to fifty dimensions, followed by t-SNE, will suppress some noise and speed up the computation of pairwise distances between samples (scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE, accessed on 26 September 2022).

3.2.4. Clustering Techniques

Clustering is a decomposition of an entity set into "natural groups" in which these groups capture the natural structure of the data. There are two significant points to clustering, the first being the algorithmic issues on how to find such data decomposition and the second being the quality of the computed decomposition [59]. Initially introduced in data mining research as an unsupervised classification method to transform patterns into groups [59]. Later, it was expanded into other fields, such as information retrieval [60] and text summarization [61]. Its concern is to group a set of entities that are similar to each other and dissimilar from entities that belong to other groups [62]. In the case of intra-cluster density versus inter-cluster sparsity [59], the objective is to minimize intra-cluster distances and maximize inter-cluster distances. Other paradigms exist, such as the density-based paradigm, which is similar to human perception, since we are used to grouping things into categories in our daily life [59].

K-means is the most known clustering technique widely used in multiple fields. It is a partitional type of clustering published by Forgy [63], and then a more efficient version was proposed and published by Hartigan [64]. This method works with a distance function between data points to decide the number of clusters needed (k). Since this technique depends on the selection of the initial centroids for its results and it is a hard clustering technique (one data point can only be in one cluster), in some cases, this can be seen as a problem. However, it is an effective technique used widely in multiple fields, becoming an excellent all-around clustering technique [65]. We used this technique in a 100-run experiment, with different centroid seeds, as a baseline technique.

K-means++ was created by Arthur and Vassilvitskii [66]. The goal is to disperse the initial centroid by assigning the first centroid randomly and then choosing the rest of the centroids based on the maximum squared distance, pushing the centroids as far as possible from one another [65]. We used this technique in a 100-run experiment with different centroid seeds.

Ckmeans, Consensus K-Means, was created by Monti et al. [67]. It consists of an unsupervised ensemble clustering algorithm, combining multiple K-Means clustering executions. Each K-Means is trained on a random subset of the data and a random subset of the features. The predicted cluster memberships of each single clustering execution are then combined into a consensus matrix, determining the number of times each pair of samples was clustered over all clustering execution [67]. We used this technique in a 100-run experiment with different centroid seeds drawing 92% of the samples and 92% of features for each run. These last two values were obtained by performing preliminary testing.

According to Sonagara and Badheka [68], hierarchical clustering involves building a cluster hierarchy using a tree of clusters, commonly known as a dendrogram. There are two basic approaches to hierarchical clustering:

- Agglomerative—Understood as a bottom-up approach, it begins with points as individual clusters and, at every step, merges the most similar or nearest pair of clusters, needing a definition of cluster similarity or distance.
- Divisive—Understood as a top-down approach, it begins with one cluster gathering all the data. At every step, it splits the cluster until singleton clusters of individual points stay, needing, at every step, a decision on which cluster to separate and how to perform the split.

We utilized the agglomerative hierarchical clustering technique with linkage equaling average, since it was the best linkage method in terms of performance in preliminary tests and backed by state-of-the-art research.

3.3. Proposed Approach

We tested several combinations of techniques and analyzed the impact on the results. Different embedders were tested considering the context (or not), different clustering techniques (partitional and hierarchical-based), and dimension reduction techniques. Their impact on metrics was analyzed. Figure 1 illustrates the pipeline we developed to run these experiments and Table 1 presents a brief overview of the used combinations for the experiments.

The same preprocessing steps were used in every approach testing. They consisted of applying the intra-phrase segmentation algorithm, followed by lowercasing, tokenization, lemmatization, removal of punctuation, and stop words. Whenever additional input data were sent to the clustering method, the min–max method was used to normalize the categorical classes.

The models' outputs will not be changed in terms of dimension reduction. Using that as a baseline: from 200 to 100 dimensions in steps of 50, from 100 to 25 in steps of 25, from 25 to 5 in steps of 5, and from 5 to 1 in steps of 1. In terms of datasets, the baseball dataset, since it is more recent, will be used as the primary dataset to test approaches. In contrast, the Cars dataset will be used as a validation dataset.

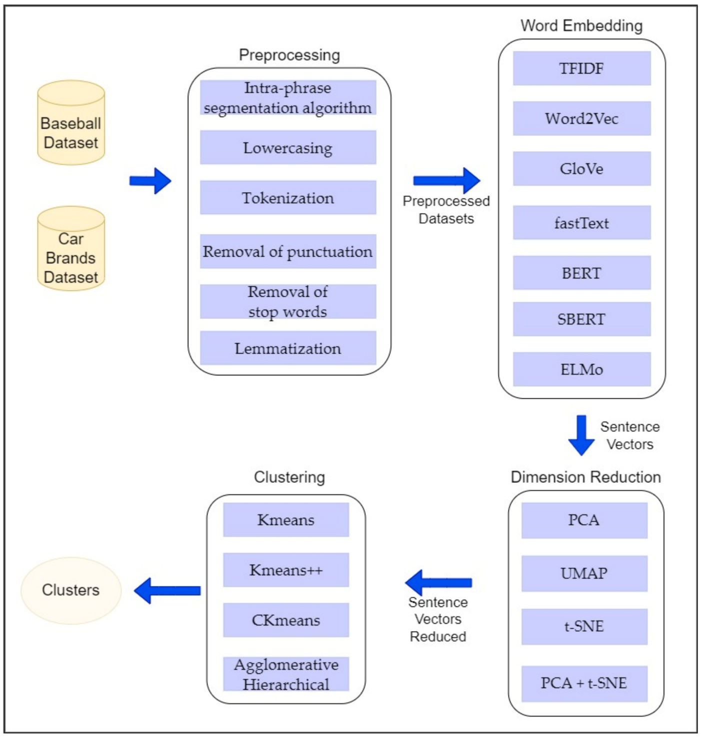

Figure 1. Methodology overview, adapted from [69].

Table 1. Quick visualization of the approach's definition.

	Technique	Implementation	Parameters
Word Embedding	TF-IDF	scikit-learn	ngram_range = (1,1)
	Word2Vec	word2vec-google-news-300	Binary mode = True max length = 200
	GloVe	glove.6B.200d	max length = 200
	fastText	crawl-300d-2M	max length = 200
	BERT	12/768 (BERT-Base)	-
	SBERT	all-MiniLM-L6-v2	-
	ELMo	v3	-
Dimensionality Reduction	PCA	scikit-learn	No change
	t-SNE	scikit-learn	200 to 100 in steps of 50
	PCA + t-SNE	PCA to 50 dimensions and then t-SNE application	100 to 25 in steps of 25 25 to 5 in steps of 5
	UMAP	umap-learn	5 to 1 in steps of 1
Clustering	Kmeans	scikit-learn	n_init = 100
	Kmeans++	scikit-learn	n_init = 100
	Agglomerative Hierarchical	scikit-learn	linkage = average
	CKmeans	pyckmeans	n_rep = 100 p_samp = 0.92 p_feat = 0.92

4. Experiments

In this section, the utilized metrics and the obtained experiment results are exposed to allow replication and further discussion in the article.

4.1. Metrics

The results of a clustering technique can be evaluated through metrics that might consider the ground truth labels if they are available. Ground truth labels are humanly provided classifications of the data on which the algorithms are trained or against which they are evaluated [70]. Some metrics will be addressed ahead.

4.1.1. Intrinsic Metrics

When ground truth labels are unavailable, only a few metrics are available to evaluate the performance of a clustering technique [71]. Silhouette is a method that provides a concise measure of how similar an object is to its cluster, compared to other clusters through the usage of distance metrics. Any metric can be used; typically, Euclidian is used to calculate the silhouette coefficient, and the results range from -1 to 1, where a high value means the clusters are well separated, minimizing the distance intra-cluster and maximizing the distance inter-cluster [72].

4.1.2. Extrinsic Metrics

When ground truth labels are available, some metrics exist to evaluate the performance of a clustering technique [71]. Mutual information functions are based on entropy, and entropy decreases as the uncertainty decreases. This way, mutual information reduces the entropy of class labels when we are given the cluster labels, allowing us to know how much the uncertainty about class labels decreases when we know the cluster labels, being similar to the information gathered in decision trees [73].

4.2. Sentences as Only Input Data

In this subset of experiments, only the sentences typed by the participants of the Reddit discussion were used as input data for each clustering technique. Initially, to evaluate which word embedders we should use moving forward, we decided to fixate the k (number of clusters hyperparameter) to the number of alternatives on the baseball dataset. Since the baseball dataset was manually annotated, the ground truth labels were available, allowing us to use the mutual information (MI) metric to evaluate the performance of the approaches. We decided to use MI, since other available metrics that use ground truth labels, such as homogeneity and completeness, have MI as part of their calculations. Furthermore, MI compares the ideal clustering results through the ground truth labels and the obtained clustering results and determines how similar both are.

4.2.1. Word Embedders Variation

As we can see in Figure 2, using K-means as a baseline clustering technique, Word2Vec, fastText, and GloVe all had similar results. Since all of them are static word vectors, GloVe was decided to be used from those three, since it was the fastest computationally wise. SBERT performed better than BERT and was much better computational-wise, hence, why BERT embeddings were dropped moving forward. Furthermore, ELMo is outperformed by SBERT as well, and since both embedders consider the context, SBERT was chosen to move forward as that type of embedder. This way, from these preliminary experiments, TF-IDF, GloVe, and SBERT were the chosen embedders to be used in the following experiments.

4.2.2. Dimensionality Reduction Techniques Variation

The process of converting raw text into word vectors used in clustering techniques produces vectors of variable size. Depending on the size of input data and the word embedding technique applied, the output vectors can reach a size in the order of the hundreds or even thousands per sentence, a length of 200 for the case of GloVe, and a length of 32,768 when we applied SBERT. This high number of features per sentence requires very high processing conditions in terms of RAM memory and processing cores, which are sometimes impossible to have, becoming computationally inefficient generally. In clustering, this is intensified because it makes it harder for a clustering technique to find similarities in the data to cluster them together when this vast number of features are supplied as input.

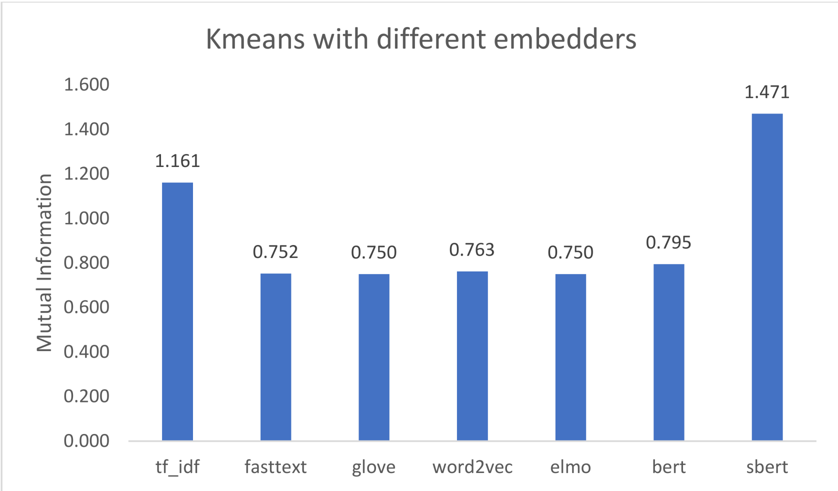

Figure 2. Kmeans performance with different word embedders on the baseball dataset.

Dimensionality reduction techniques, as stated in Section 3.2.2, enhance the methodology's computational efficiency and improve the clustering techniques' performance. Despite the known benefits of dimension reduction techniques, some disadvantages may come with their application. For instance, reducing features may lead to information loss. In addition to that, additional effort is needed to run several tests to find the optimal value for dimensionality reduction.

Having decided on word embeddings to be used in the experiments, we chose not to fixate the number of clusters (K) and test it out with multiple K ranging from 2 to 128 in exponentials of 2. With this change, our ground labels could not be used, since the number of clusters might not be the same as the number of unique ground labels, leading to only the silhouette score getting used. The combination of the silhouette score and the K value it maxes out for each approach will dictate the quality of the results.

Figure 3 shows the best silhouette score for each technique without applying dimensionality reduction techniques. In contrast, Figure 4 shows how the clustering results improve when putting all the embedders with the exact final dimensions, which are 200 from the GloVe word embedder technique. We can see a tendency where UMAP outperforms PCA in this number of dimensions.

Figure 5 shows how the silhouette score varies when reducing the number of dimensions of the embeddings; analyzing the tendencies of the approaches, since multiple approaches are overlapping, making it less readable, we can see that most approaches presented there show a minimal increase in performance until ten dimensions, where it starts to steadily increase as dimensions get reduced, except the agglomerative hierarchical clustering technique with GloVe word embedder and PCA dimensionality reduction that shows a stabilization until 15 dimensions and then a decrease in performance and joining the tendency of the remaining approaches on that graph after ten dimensions. The same pattern can be seen in Figure 6 with the TFIDF with PCA, which spiked in performance after ten dimensions. The GloVe with UMAP does not have a perceivable pattern where it spikes at specific dimensions. Based on most approach patterns, a decision was made to start experimenting from only ten until one dimension.

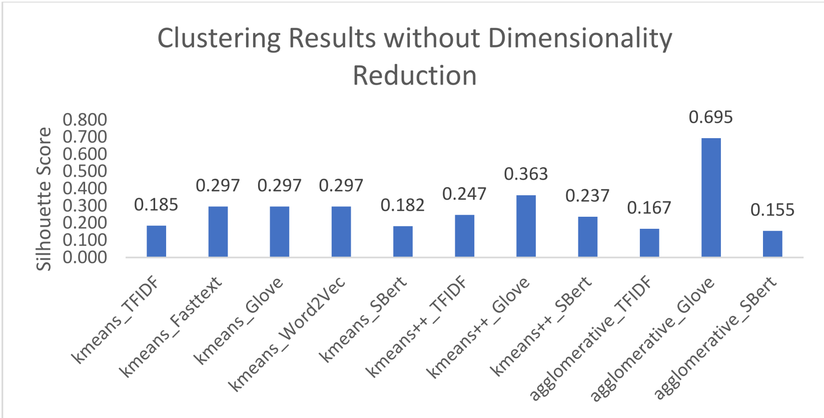

Figure 3. Clustering results in the new test settings on the baseball dataset.

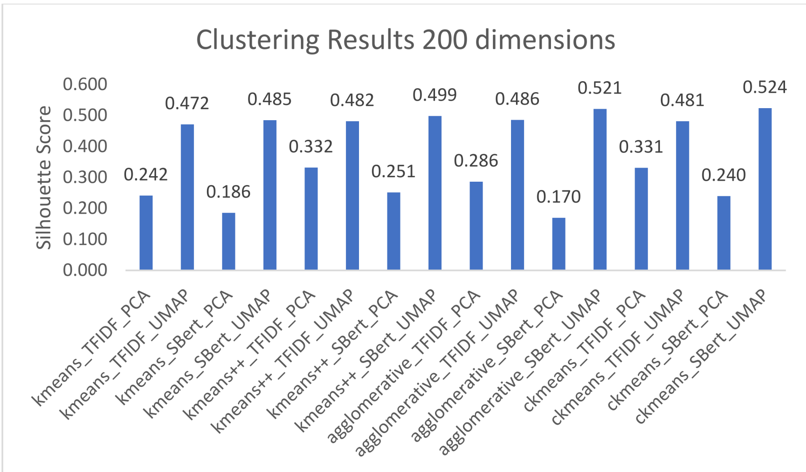

Figure 4. Clustering results with every approach at the 200 dimensions on the baseball dataset.

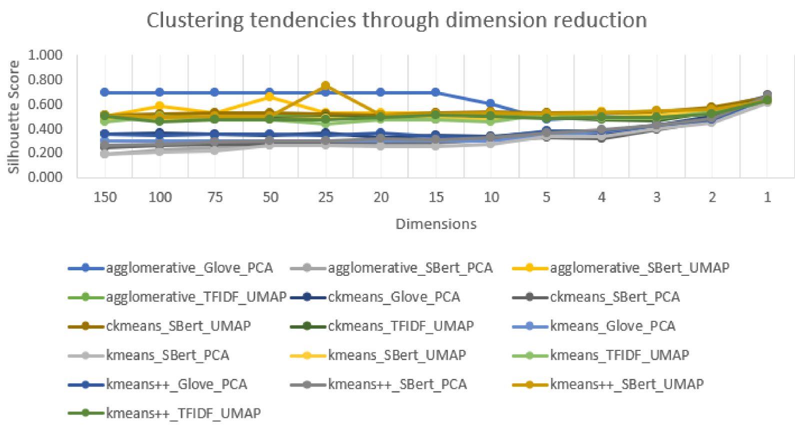

Figure 5. Clustering tendencies through dimension reduction on the baseball dataset.

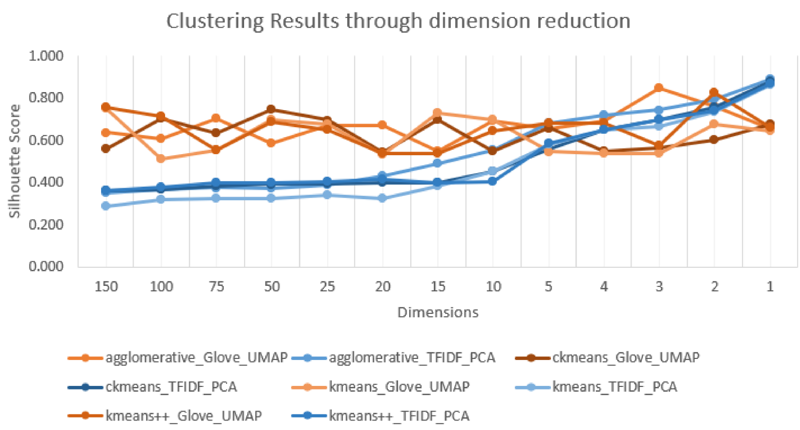

Figure 6. Clustering results through dimension reduction on the baseball dataset.

Even though it displays promising results silhouette score-wise, when analyzing the number of clusters used to reach those values, they either maximize at the 128 clusters (Figure 7A) or on 2 clusters (Figure 7B), which is not the expected result for this task, since 2 clusters group the data with no perceivable differences and 128 clusters are too many clusters with no interest for the solution of the problem that this works intends to solve. Therefore, a tradeoff between the silhouette score and the number of clusters at which a specific approach maxed its silhouette score is considered when evaluating the obtained results.

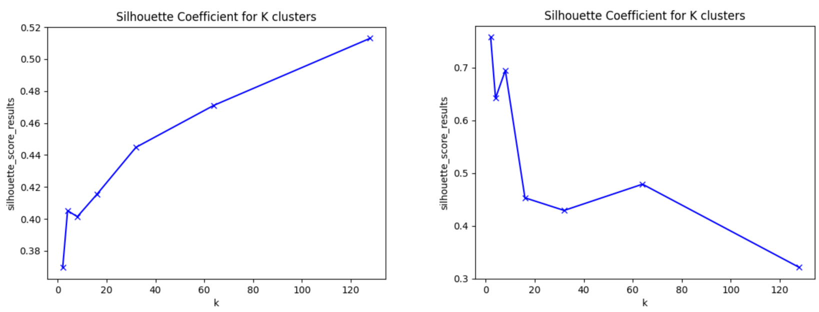

Figure 7. Example of approaches performance through multiple Ks. (**A**) Maximizing in 128 clusters, (**B**) Maximizing in 2 clusters.

4.3. Addition of Polarity to the Input Data

Adding Polarity to the input data did not change the results silhouette score-wise nor on a brief analysis of the resulting clusters. Still, the number of clusters for the best silhouette score of each approach starts to show good values, with some methods maximizing at four and eight clusters, each being more in line with the expected number of clusters when they are not predefined. Since one sentence can have multiple annotations, brief experiments were done to minimize the number of duplicated sentences, no duplicates, and two dupes max, with no interesting results to appoint.

4.4. Addition of Alternative and Criterion to the Input Data

Adding alternative and criterion to the input data (which was already sentenced and polarity) provided some exciting results with one approach that maximized at four clusters showing that the kmeans++ clustering technique with TFIDF word embedder and the

PCA reducing to two dimensions divided the data into polarity and criterion. Whether the criterion part existed or not, it created clusters with positive polarity with criterion and created negative polarity with criterion, etc. Unfortunately, this was not the objective of the work. Still, it was interesting that an unsupervised technique made such a division, which led us to believe that the method put too much emphasis on those two features and was not equally distributed. Only adding alternative and criteria to the sentence without polarity did not achieve any exciting results.

4.5. Addition of Alternative to the Implicit Sentences Text

This way, we believed that going entirely for the sentence as the only input for the clustering technique would bring more desirable results. With just the input sentence, we adjusted some parameters, such as the range of K to be from 2 to 20 in steps of 1 and decided to add the alternative to the input sentence, resulting in high-quality clusters closely related to what was expected. We found that two dimensions was the best amount for reducing dimensions, since it is more in line with practices of the area where a reduction to two dimensions for visualization is made. Most approaches maximized their silhouette score at a sufficient number of K (not in the lowest value of 2 or the highest value of 20) with two dimensions. The best approaches did not show any improvements between the two and one dimensions.

Adding the alternative at the beginning or the end of the sentence did not wield any significant changes to the silhouette score.

5. Discussion

After obtaining the experiment's results, understanding and evaluating them is required to develop an approach to solve the problem.

As we can see in Figure 8, the best approach was the agglomerative hierarchical clustering technique with TFIDF word embedder and PCA reducing to two dimensions. However, this approach reached the best value at two clusters, which we previously discarded; since the goal is to organize conversations, splitting them into meaningful groups, only two clusters would be too reductive. Therefore, the accepted approach to solve the objective of this work was the kmeans++ clustering technique with SBERT word embedder and UMAP reducing to two dimensions, which resulted in eight clusters.

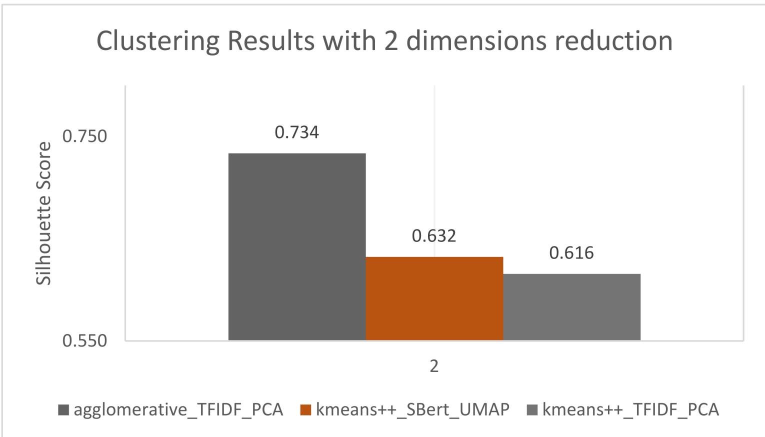

Figure 8. Best clustering results with two dimensions reduction on the baseball dataset.

Analyzing these clusters manually and through word clouds (Figure 9A–H) made us accept this approach because of the variety and the excellent division between them, even though it has a silhouette score of 0.63.

Figure 9. Word clouds with the clustering results of the baseball dataset.

Validating this approach on the other dataset we had, the car brands dataset, wielded good results but not as good as the baseball dataset results, with 16 clusters and 0.59 silhouette score. The word clouds (Figures 10A–H and 11A–H) show some variety, but some clusters could have been better divided, since they refer to different topics.

Figure 10. Word clouds with some clusters results of the car brands dataset.

This discrepancy could be attributed to multiple factors, such as the annotation quality of the datasets, their size (since the baseball dataset is bigger than the cars dataset), the distribution of alternatives in each dataset (number of times they appear in messages), the quality of the discussion, and the arguments used. Nevertheless, we believe this approach can achieve the objective of dynamically organizing the conversation based on the arguments used. In a real setting, not fixating on the value of clusters and with even more data, this approach manages to pick the correct number of clusters for the input data and group that data into perceivable clusters that can then be utilized in intelligent reports for the decision makers.

Figure 11. Word clouds with the remaining clusters results of the car brands dataset.

6. Conclusions

This work aimed to study the application of clustering techniques to the context of group decision making to dynamically generate clusters of the messages exchanged by decision makers.

To achieve the proposed goals, we studied and experimented with multiple clustering techniques, word embedders, and dimensionality reduction techniques to understand what configuration of these three techniques works best for the GDM context. From the tested experiments, the best approach consisted of applying the K-means++ clustering technique with SBERT word embedder and UMAP dimensionality reduction technique, reducing to two dimensions, which resulted in eight clusters with 0.63 silhouette score. Using the same approach on the validation dataset (car brands dataset) obtained satisfactory results but not as good as in the baseball dataset. This difference in results can be related to the small dimension of the car brands dataset and its higher dispersion concerning alternatives, leading to a higher number of formed clusters containing few observations. However, we believe this approach is a feasible solution for the problem we intend to tackle. In a real environment, this approach will automatically pick the correct number of clusters and group the data into perceivable clusters that can then be utilized in intelligent reports for decision makers.

This work contributed to the enhancement of our GDSS prototype, enabling a new feature capable of presenting clusters of messages to decision makers inside the intelligent reports. With this feature, intelligent reports for GDSS become even more helpful, since it can dynamically organize the conversations taking place by grouping them on the arguments used, allowing the decision makers to have a better perception of the direction of the conversation.

In future work, we intend to continue using these two datasets for other experiments, such as a model to detect criteria used in an argument or sentiment analysis models to predict the polarity of the argument in a discussion. Furthermore, the inclusion of this work on the fully fledged GDSS with strong AM models is planned.

Author Contributions: L.C.: Investigation, Conceptualization, Validation, Writing—review and editing. V.R.: Investigation, Software, Writing—original draft. J.M.: Validation, Writing—review and editing. G.M.: Supervision, Writing—review and editing. P.N.: Supervision, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Funds through the Portuguese FCT—Fundação para a Ciência e a Tecnologia under the R&D Units Project Scope UIDB/00319/2020, UIDB/00760/2020, UIDP/00760/2020, and by the Luís Conceição Ph.D. Grant with the reference SFRH/BD/137150/2018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Datasets used in this work are available at https://cloud.gecad.isep.ipp.pt/s/mDMPSPbrAQ9jFge.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Liu, W.; Dong, Y.; Chiclana, F.; Cabrerizo, F.J.; Herrera-Viedma, E. Group decision-making based on heterogeneous preference relations with self-confidence. *Fuzzy Optim. Decis. Mak.* **2016**, *16*, 429–447. [CrossRef]
2. Carneiro, J.; Alves, P.; Marreiros, G.; Novais, P. Group decision support systems for current times: Overcoming the challenges of dispersed group decision-making. *Neurocomputing* **2020**, *423*, 735–746. [CrossRef]
3. Carneiro, J.; Martinho, D.; Alves, P.; Conceição, L.; Marreiros, G.; Novais, P. A Multiple Criteria Decision Analysis Framework for Dispersed Group Decision-Making Contexts. *Appl. Sci.* **2020**, *10*, 4614. [CrossRef]
4. Kaner, S. *Facilitator's Guide to Participatory Decision-Making*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2014. Available online: https://www.wiley.com/en-us/Facilitator%27s+Guide+to+Participatory+Decision+Making%2C+3rd+Edition-p-9781118404959 (accessed on 18 March 2022).
5. Carneiro, J.; Martinho, D.; Marreiros, G.; Novais, P. Arguing with Behavior Influence: A Model for Web-Based Group Decision Support Systems. *Int. J. Inf. Technol. Decis. Mak.* **2019**, *18*, 517–553. [CrossRef]
6. Chen, C.-T. Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets Syst.* **2000**, *114*, 1–9. [CrossRef]

7. Majumder, M. Multi Criteria Decision Making. In *Impact of Urbanization on Water Shortage in Face of Climatic Aberrations*; Springer: Singapore, 2015; pp. 35–47. [CrossRef]
8. Carneiro, J.; Martinho, D.; Marreiros, G.; Jimenez, A.; Novais, P. Dynamic argumentation in UbiGDSS. *Knowl. Inf. Syst.* **2017**, *55*, 633–669. [CrossRef]
9. Tang, M.; Liao, H. From conventional group decision making to large-scale group decision making: What are the challenges and how to meet them in big data era? A state-of-the-art survey. *Omega* **2019**, *100*, 102141. [CrossRef]
10. Conceição, L.; Martinho, D.; Andrade, R.; Carneiro, J.; Martins, C.; Marreiros, G.; Novais, P. A web-based group decision support system for multicriteria problems. *Concurr. Comput. Pr. Exp.* **2019**, *33*, 499–534. [CrossRef]
11. Lawrence, J.; Reed, C. Argument Mining: A Survey. *Comput. Linguistics* **2020**, *45*, 765–818. [CrossRef]
12. Zuheros, C.; Martínez-Cámara, E.; Herrera-Viedma, E.; Herrera, F. Sentiment Analysis based Multi-Person Multi-criteria Decision Making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews. *Inf. Fusion* **2020**, *68*, 22–36. [CrossRef]
13. Lippi, M.; Torroni, P. Argument Mining: A Machine Learning Perspective. *Lect. Notes Comput. Sci.* **2015**, *9524*, 163–176. [CrossRef]
14. Conceição, L.; Carneiro, J.; Martinho, D.; Marreiros, G.; Novais, P. Generation of Intelligent Reports for Ubiquitous Group Decision Support Systems. In Proceedings of the 2016 Global Information Infrastructure and Networking Symposium, Porto, Portugal, 19–21 October 2016; pp. 1–6. [CrossRef]
15. Kim, C.; Yin, P.; Soto, C.; Blaby, I.; Yoo, S. Multimodal Biological Analysis Using NLP and Expression Profile. In Proceedings of the 2018 New York Scientific Data Summit (NYSDS), New York, NY, USA, 6–8 August 2018. [CrossRef]
16. Sarkar, S.; Lodhi, V.; Maiti, J. Text-Clustering Based Deep Neural Network for Prediction of Occupational Accident Risk: A Case Study. In Proceedings of the 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 15–17 November 2018. [CrossRef]
17. Hema, D.; David, V. Fuzzy Clustering and Genetic Algorithm for Clinical Pratice Guideline Execution Engines. In Proceedings of the 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, India, 14–16 December 2017; pp. 13–16. [CrossRef]
18. Dragos, D.; Schmeelk, S. What are they Reporting? Examining Student Cybersecurity Course Surveys through the Lens of Machine Learning. In Proceedings of the 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 14–17 December 2020; pp. 961–964. [CrossRef]
19. Gupta, A.; Tripathy, B. Implementing GloVe for Context Based k-Means++ Clustering. In Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017, Palladam, India, 7–8 December 2017; pp. 1041–1046. [CrossRef]
20. Huang, Y.; Chen, C.; Xing, Z.; Lin, T.; Liu, Y. Tell them apart: Distilling technology differences from crowd-scale comparison discussions. In Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, Montpellier, France, 3–7 September 2018; pp. 214–224. [CrossRef]
21. Liu, Y.; Whitfield, C.; Zhang, T.; Hauser, A.; Reynolds, T.; Anwar, M. Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning. *Health Inf. Sci. Syst.* **2021**, *9*, 1–16. [CrossRef]
22. Reimers, N.; Schiller, B.; Beck, T.; Daxenberger, J.; Stab, C.; Gurevych, I. Classification and clustering of arguments with contextualized word embeddings. In Proceedings of the ACL 2019—57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 567–578. [CrossRef]
23. Färber, M.; Steyer, A. Towards Full-Fledged Argument Search: A Framework for Extracting and Clustering Arguments from Unstructured Text. 2021. Available online: http://arxiv.org/abs/2112.00160 (accessed on 28 September 2022).
24. Dumani, L.; Schenkel, R. Quality-Aware Ranking of Arguments. In Proceedings of the 29th ACM International Conference on Information and Knowledge, New York, NY, USA, 19–23 October 2022; pp. 335–344. [CrossRef]
25. Dumani, L.; Neumann, P.J.; Schenkel, R. A Framework for Argument Retrieval. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; Volume 12035, pp. 431–445. [CrossRef]
26. Daxenberger, J.; Schiller, B.; Stahlhut, C.; Kaiser, E.; Gurevych, I. ArgumenText: Argument Classification and Clustering in a Generalized Search Scenario. *Datenbank Spektrum* **2020**, *20*, 115–121. [CrossRef]
27. Cardoso, T.; Rodrigues, V.; Conceição, L.; Carneiro, J.; Marreiros, G.; Novais, P. Aspect Based Sentiment Analysis Annotation Methodology for Group Decision Making Problems: An Insight on the Baseball Domain. In *Information Systems and Technologies*; Springer: Cham, Switzerland, 2022; pp. 25–36.
28. Denny, M.J.; Spirling, A. Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Politi. Anal.* **2018**, *26*, 168–189. [CrossRef]
29. Weiss, S.; Indurkhya, N.; Zhang, T.; Damerau, F. *Text Mining: Predictive Methods for Analyzing Unstructured Information*; Springer: New York, NY, USA, 2004; pp. 1–237. [CrossRef]
30. Ramasubramanian, C.; Ramya, R. Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm. *Int. J. Adv. Res. Comput. Commun. Eng.* **2013**, *2*, 4536–4538. Available online: www.ijarcce.com (accessed on 28 September 2022).
31. Čibej, J.; Fišer, D.; Erjavec, T. Normalisation, Tokenisation and Sentence Segmentation of Slovene Tweets. In *Normalisation and Analysis of Social Media Texts (NormSoMe)*; 2016; Volume 2016, pp. 5–10. Available online: http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-NormSoMe_Proceedings.pdf#page=10 (accessed on 28 September 2022).

32. Wicks, R.; Post, M. A unified approach to sentence segmentation of punctuated text in many languages. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Bangkok, Thailand, 1–6 August 2021; pp. 3995–4007. [CrossRef]
33. Ganesan, K. All you need to know about Text Preprocessing for Machine Learning & NLP. KDnuggets 2019. Available online: https://www.kdnuggets.com/2019/04/text-preprocessing-nlp-machine-learning.html (accessed on 27 January 2022).
34. Kumar, S. Getting started with Text Preprocessing. Kaggle. 2019. Available online: https://www.kaggle.com/sudalairajkumar/getting-started-with-text-preprocessing/notebook (accessed on 27 January 2022).
35. Goldberg, Y. Neural Network Methods for Natural Language Processing. *Synth. Lect. Hum. Lang. Technol.* **2017**, *10*, 92. [CrossRef]
36. Brownlee, J. Machine Learning Mastery: What Are Word Embeddings for Text? *Mach. Learn. Mastery* **2019**. Available online: https://machinelearningmastery.com/what-are-word-embeddings/ (accessed on 11 February 2022).
37. Leskovec, J.; Rajaraman, A.; Ullman, J. *Mining of Massive Datasets*; Cambridge University Press: Cambridge, UK, 2014. [CrossRef]
38. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed representations of words and phrases and their compositionality. *Adv. Neural. Inf. Process. Syst.* **2013**, *2*, 3111–3119.
39. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, AZ, USA, 2–4 May 2013; pp. 1–12.
40. Karani, D. Introduction to Word Embedding and Word2Vec. *Towards Data Sci.* **2018**. Available online: https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa (accessed on 11 February 2022).
41. Pennington, J.; Socher, R.; Manning, C. Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543. [CrossRef]
42. Theiler, S. Basics of Using Pre-trained GloVe Vectors in Python. *Medium. Anal. Vidhya* **2019**. Available online: https://medium.com/analytics-vidhya/basics-of-using-pre-trained-glove-vectors-in-python-d38905f356db#:~{}:text=GlobalVectorsforWordRepresentation,inahigh-dimensionalspace (accessed on 11 February 2022).
43. Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. Enriching Word Vectors with Subword Information. *Trans. Assoc. Comput. Linguistics* **2017**, *5*, 135–146. [CrossRef]
44. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the NAACL HLT 2019–2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Volume 1, pp. 4171–4186. [CrossRef]
45. McCormick, C.; Ryan, N. BERT Word Embeddings Tutorial. Available online: http://www.mccormickml.com (accessed on 16 August 2022).
46. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 3980–3990. [CrossRef]
47. Peters, M.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep contextualized word representations. In Proceedings of the NAACL HLT 2018—2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, LA, USA, 1–6 June 2018; Volume 1, pp. 2227–2237. [CrossRef]
48. Cunningham, P. Dimension reduction. In *Machine Learning Techniques for Multimedia*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 91–112. [CrossRef]
49. Allaoui, M.; Kherfi, M.; Cheriet, A. Considerably improving clustering algorithms using umap dimensionality reduction technique: A comparative study. *Lect. Notes Comput. Sci.* **2020**, *12119*, 317–325. [CrossRef]
50. Pearson, K. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
51. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [CrossRef]
52. Hotelling, H. Relations Between Two Sets of Variates. *Biometrika* **1936**, *28*, 321. [CrossRef]
53. Kumar, A. Principal Component Analysis with Python. GeeksforGeeks. 2022. Available online: https://www.geeksforgeeks.org/principal-component-analysis-with-python/ (accessed on 17 August 2022).
54. Kambhatla, N.; Leen, T.K. Dimension Reduction by Local Principal Component Analysis. *Neural Comput.* **1997**, *9*, 1493–1516. [CrossRef]
55. Stewart, G.W. On the Early History of the Singular Value Decomposition. *SIAM Rev.* **1993**, *35*, 551–566. [CrossRef]
56. Hinton, G.; Roweis, S. Stochastic Neighbor Embedding. *Adv. Neural Inf. Process. Syst.* **2002**, *15*, 857–864. Available online: https://papers.nips.cc/paper/2002/file/6150ccc6069bea6b5716254057a194ef-Paper.pdf (accessed on 17 August 2022).
57. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
58. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 851. [CrossRef]
59. Gaertler, M. Clustering. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Brandes, U., Erlebach, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3418, pp. 178–215. [CrossRef]
60. Bordogna, G.; Pasi, G. Soft clustering for information retrieval applications. *WIREs Data Min. Knowl. Discov.* **2011**, *1*, 138–146. [CrossRef]

61. Deshpande, A.; Lobo, L. Text Summarization using Clustering Technique. *Int. J. Eng. Trends Technol.* **2013**, *4*, 8. Available online: http://julio.staff.ipb.ac.id/files/2011/12/IJETT-V4I8P120.pdf (accessed on 28 September 2022).
62. Bramer, M. Clustering. In *Principles of Data Mining*; Springer: London, UK, 2007; pp. 221–238. [CrossRef]
63. Forgy, E. Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics* **1965**, *21*, 768–769.
64. Hartigan, J. Clustering Algorithms. *J. Mark. Res.* **1997**, *14*, 1. [CrossRef]
65. Iliassich, L. Clustering Algorithms: From Start To State Of The Art. Toptal. 2016. Available online: https://www.toptal.com/machine-learning/clustering-algorithms (accessed on 10 February 2022).
66. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms* **2007**, *7–9*, 1027–1035. [CrossRef]
67. Monti, S.; Tamayo, P.; Mesirov, J.; Golub, T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Mach. Learn.* **2003**, *52*, 91–118. [CrossRef]
68. Sonagara, D.; Badheka, S. Comparison of Basic Clustering Algorithms. *Int. J. Comput. Sci. Mob. Comput.* **2014**, *3*, 58–61.
69. Kang, Y.; Cai, Z.; Tan, C.-W.; Huang, Q.; Liu, H. Natural language processing (NLP) in management research: A literature review. *J. Manag. Anal.* **2020**, *7*, 139–172. [CrossRef]
70. Urumov, G. The Solid Facts of Ground Truth Annotations. UNDERSTAND.AI. 2021. Available online: https://understand.ai/blog/annotation/machine-learning/autonomous-driving/2021/05/31/ground-truth-annotations.html (accessed on 9 September 2022).
71. Han, J.; Kamber, M.; Pei, J. Cluster Analysis. In *Data Mining*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 443–495. [CrossRef]
72. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
73. Yıldırım, S. Evaluation Metrics for Clustering Models. *Towards Data Sci.* **2021**. Available online: https://towardsdatascience.com/evaluation-metrics-for-clustering-models-5dde821dd6cd (accessed on 9 September 2022).

Chapter 3

Conclusions

The last chapter, chapter 3, presents the resulting contributions from the works described in this thesis and also how they validate the hypothesis previously defined in the chapter 1. The activities developed during the period of this Ph.D. are also described in the remainder of this chapter. Section 3.2 presents Collaboration in Research Projects subsection 3.2.1, other publications (subsection 3.2.2) that were made during this period related to other scientific research projects but that are somehow related to the core topic of this thesis; collaboration and participation in scientific reviews and events are described in subsection 3.2.3; Lecturing activities carried out during this Ph.D. period subsection 3.2.4; students supervision and the work they developed are presented in subsection 3.2.5. Finally, in section 3.3, some final remarks and considerations about the developed work are presented, as well as some guidelines for future work.

3.1 Contributions

This Ph.D. thesis was elaborated as a set of publications that encompass significant contributions in using machine learning algorithms to process argumentation dialogues in the context of group decision-making. The accomplishment of each one of the objectives described in Section 1.5 is demonstrated in Table 2, by matching each objective with the corresponding solution in a section of Chapter 2.

Table 2: List of the defined objectives and respective sections where they were addressed.

Objective	Section
Objective 1: Study and develop a model that allows the system itself to assess the arguments introduced by real participants automatically.	section 2.2 and section 2.3
Objective 2: Study and develop models that will be able to dynamically generate clusters of the messages exchanged by decision-makers.	section 2.3 and section 2.4
Objective 3: Design a web-based GDSS prototype based on a multi-agent system that will include the models previously described.	section 2.1

To validate objective 1 - Study and develop a model that allows the system itself to assess the arguments introduced by real participants automatically, ML supervised classification algorithms were applied to a dataset containing arguments extracted from online debates, creating an argumentative sentence classifier that can perform automatic classification (InFavour/Against) of argumentative sentences exchanged in GDSS's by decision-makers.In a distributed scenario where participants, which can be thousands, are supported/represented by intelligent agents, this model is very relevant, as it allows systematizing the information that is embedded in argumentative dialogues.

The objective 2 - Study and develop models that will be able to dynamically generate clusters of the messages exchanged by decision-makers, was validated through the application of clustering techniques to a dataset created based on discussions acquired from a subreddit dedicated to Baseball, with these discussions focusing on the best Baseball player ever. The annotation of the dataset and the established methodology enabled an approach to the problem of turning unstructured discussions into structured annotated data, which is needed for the Group Decision-Making (GDM) domain. To this dataset unsupervised techniques, were applied in order to group arguments on topics around a specific alternative, for example, or a discussion comparing two alternatives. This approach allowed the achievement of the goal of dynamically organizing the conversation based on the arguments used. With this model, intelligent reports for GDSS [21] become even more helpful since they can dynamically organize the conversations between decision-makers by grouping them based on, for instance, the alternatives, the criteria, and comparisons, among others, that are being discussed.

The last and third objective is - Design a web-based GDSS prototype based on a multi-agent system that will include the models previously described, was achieved by the proposal of a web-based GDSS that is able to support groups of decision-makers regardless of their geographic location. The system allows the creation of multi-criteria problems and the configuration of the preferences, intentions, and interests of each decision-maker. This web-based GDSS includes the intelligent reports dashboards that are supported by the work achieved in objectives one and two. An initial demo of this prototype, entitled "A consensus-based group decision support system using a multi-agent MicroServices approach"was presented at the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS'20).

The achievement of each of the previously identified objectives allowed the validation of the research questions identified in section 1.4 and thus responding positively to the hypothesis defined at the beginning of this work, that is, that the incorporation of Machine Learning techniques to support argumentation dialogues in web-based Group Decision Support Systems improves this kind of systems and consequently the quality of decision process.

3.2 Dissemination of Results

The work developed along this Ph.D. project was also related to the research project GrouPlanner, which goal was to conduct studies under the topic of Group Recommender Systems. This project will be described below in the first subsection (subsection 3.2.1). Dissemination of results has been made through different formats; besides collaboration in research projects, other publications have been done, participation in international events, lecturing, and supervision of students at Instituto Superior de Engenharia do Porto (ISEP).

3.2.1 Collaboration in Research Projects

The work developed during this Ph.D. project was also related to some ongoing or finished research projects. A short description of each one is presented below. Within the scope of the Ph.D., the possibility of applying some of the data analysis techniques, namely to a specific domain, of healthcare, particularly in vascular diseases, began to be studied in project Inno4Health which will be described. In the next section 3.2.2, some publications that resulted from these collaborations.

- **GrouPlanner - Supporting Groups in Travel Planning Considering Social and Cultural Aspects**
  Project GrouPlanner goal was to conduct studies, research, and experimentation under the topic of Group Recommender Systems and was meant to support groups of Chinese tourists in every aspect of travel planning and selection of Point of Interest (POI) to visit in the North of Portugal. The system should, in the first instance, aid groups in the selection of one (e.g.: hotel, city, etc.) or multiple alternatives (e.g.: POI) from a set of alternatives (according to the preferences and interests of each group member) and in a following step, support each group member along the stay, instilling a greater sense of security. As a result, the system intends to maximize group satisfaction in several dimensions. The system should be as ubiquitous as possible and consider social and cultural

aspects. This project involved ISEP, Turismo do Porto e Norte de Portugal, E.R., and Institute of Policy and Management, Chinese Academy of Sciences. This project was funded by Fundação para a Ciência e a Tecnologia (FCT), under the project reference: PTDC/CCI-INF/29178/2017.

- **TheRoute - Tourism and Heritage Routes including Ambient Intelligence with Visitants' Profile Adaptation and Context Awareness**
  Project TheRoute aimed to carry out studies, research, and experimentation around the challenge of automatically generating routes for visitors to points of interest related to Tourism and Heritage in Northern Portugal. Suggestions fit the profile of visitors, and the context and are developed around a place, route, or theme. Mobility between points of interest, inherent restrictions (schedules, accessibility), and issues related to health and well-being were also considered. The project was led by the Instituto Politécnico do Porto, with Doctor Carlos Ramos as the main researcher, and also involved as partners the Instituto Superior de Engenharia do Porto, the Polytechnic Institute of Viana do Castelo, and the company DouroAzul. This project was funded by the "Programa Operacional da Região Norte"in a joint call with the FCT with the funding reference SAICT/23447.

- **PHE - Personal Health Empowerment/AirDoc – Aplicação Móvel Inteligente para Suporte Individualizado e Monitorização da Função e Sons Respiratórios de Doentes Obstrutivos Crónicos**
  The PHE project aimed to enable people to monitor and improve their health using data analysis, gamification, and coaching strategies. The international consortium is made up of Portuguese (University of Porto, MEDIDA and ISEP), Spanish (Experis ManpowerGroup, S.L.U), Belgian (KU Luven and IDEWE Group) and Turkish (ARD GROUP and MANTIS software) partners. This project was funded by the "Programa Operacional da Região Norte"with the funding reference: ANI|P2020 project nr 033275.

- **WBL_SP! - from birth to adult age - a WBL Successful Practice!**
  The WBL_SP! Project aimed to develop a digital platform for the management and validation of skills acquired within the scope of attending professional courses. The consortium is made up of partners from 3 countries, namely Portugal, Spain and Italy: AEVA – Association for the Education and Enhancement of the Region of Aveiro (PT), A. Silva Matos Metalomecânica S.A. (PT), Municipality of Sever do Vouga (PT), Politeknika Ikastegia Txorierri (ES), TKNIKA (ES), Gestamp Technology Institute (IT), ITIS "E. Mattei" (IT), Provincia di Pesaro and Urbino (IT), Benelli Armi SPA (IT) and Polytechnic Institute of Porto (PT). This was an Erasmus+ project with the reference: 2017-1-PT01-KA202-03590.

- **Inno4Health - Stimulate continuous monitoring in personal and physical health**
  Inno4Health aims to stimulate innovation in continuous health and fitness monitoring in order to inform patients and their physician on their readiness regarding surgery and the ability to recover rapidly from invasive treatment. In sports, the same technology will be used to continuously assess

fitness and health in order to provide information to athletes and their coaches and to help them optimise their performance during competitions. The international project consortium is formed by 26 partners from 6 countries, namely Portugal, Netherlands, Turkey, Lithuania, Romania and Canada. The national consortium has the participation of ISEP, Faculdade de Medicina da Universidade do Porto (FMUP) and the company Wiseware Ltd. This project is funded by Compete 2020 with the reference: POCI-01-0247-FEDER-069523.

3.2.2 Other publications

Participation and collaboration with colleagues that work in some of the previously identified research projects, resulted in the publication of the following papers:

- **Predicting satisfaction: Perceived decision quality by decision-makers in Web-based group decision support systems**
  João Carneiro, Pedro Saraiva, Luís Conceição, Ricardo Santos, Goreti Marreiros, and Paulo Novais. "Predicting satisfaction: Perceived decision quality by decision-makers in Web-based group decision support systems". In: *Neurocomputing* (2019). issn: 0925-2312. doi: `doi.org/10.1016/j.neucom.2018.05.126`
  Abstract. In future, the organizations' likelihood to endure and succeed will depend greatly on the quality of every decision made. It is known that most decisions in organizations are made in group. With the purpose of supporting decision-makers anytime and anywhere, Web-based Group Decision Support Systems (GDSS) have been studied. The amount of Web-based GDSS incorporating automatic negotiation mechanisms such as argumentation has been steadily increasing. Usually, these systems/models are evaluated through mathematical proofs, number of rounds or seconds to propose (reach) a solution. However, those techniques are not very informative in terms of the decision quality. Here, we propose a model that intends to predict the decision-makers' satisfaction (perception of the decision quality), specifically designed to deal with multi-criteria problems. Our model considers aspects such as: meeting's outcomes, decision-maker's intentions, expectations and emotional cost. To validate the proposed model in terms of its ability to predict decision-makers' satisfaction, we developed a prototype of a Web-based GDSS to be used in a case study where the participant had to make a joint decision. The decision process consisted in a set of 5 rounds, where the participant could (re)configure his/her preferences along the process. The satisfaction model ascertained its ability to predict the participants' satisfaction and allowed to understand that (as is stated in the literature) the inclusion of cognitive and emotional variables is essential to evaluate satisfaction more accurately.

- **A multiple criteria decision analysis framework for dispersed group decision-making contexts**

  João Carneiro, Diogo Martinho, Patrícia Alves, Luís Conceição, Goreti Marreiros, and Paulo Novais. "A multiple criteria decision analysis framework for dispersed group decision-making contexts". In: *Applied Sciences* 10.13 (2020), p. 4614

  Abstract. To support Group Decision-Making processes when participants are dispersed is a complex task. The biggest challenges are related to communication limitations that impede decision-makers to take advantage of the benefits associated with face-to-face Group Decision-Making processes. Several approaches that intend to aid dispersed groups attaining decisions have been applied to Group Decision Support Systems. However, strategies to support decision-makers in reasoning, understanding the reasons behind the different recommendations, and promoting the decision quality are very limited. In this work, we propose a Multiple Criteria Decision Analysis Framework that intends to overcome those limitations through a set of functionalities that can be used to support decision-makers attaining more informed, consistent, and satisfactory decisions. These functionalities are exposed through a microservice, which is part of a Consensus-Based Group Decision Support System and is used by autonomous software agents to support decision-makers according to their specific needs/interests. We concluded that the proposed framework greatly facilitates the definition of important procedures, allowing decision-makers to take advantage of deciding as a group and to understand the reasons behind the different recommendations and proposals.

- **A consensus-based group decision support system using a multi-agent MicroServices approach**

  João Carneiro, Rui Andrade, Patrícia Alves, Luís Conceição, Paulo Novais, and Goreti Marreiros. "A consensus-based group decision support system using a multi-agent MicroServices approach". In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems.* 2020, pp. 2098–2100

  Abstract. In this demo, we present a Consensus-based Group Decision Support System that makes use of a Multi-Agent Microservices approach. The proposed system comprises a Client App, an API Gateway, and a set of Microservices where different Artificial Intelligence methods are implemented. It allows to create and manage Multiple Criteria Decision Problems and to support dispersed decision-makers, allowing them to take advantage of the benefits associated with face-to-face Group Decision-Making processes.

- **A context-awareness approach to tourism and heritage routes generation**

  Carlos Ramos, Goreti Marreiros, Constantino Martins, Luiz Faria, Luís Conceição, Joss Santos, Luís Ferreira, Rodrigo Mesquita, and Lucas Schwantes Lima. "A context-awareness approach to tourism and heritage routes generation". In: *International Symposium on Ambient Intelligence*. Springer. 2018, pp. 10–23

Abstract. The aims of the TheRoute (Tourism and Heritage Routes including Ambient Intelligence with Visitants' Profile Adaptation and Context Awareness) project is to conduct studies, research and experimentation around the challenge of automatic generation of routes for visitors. The suggested routes fit the profile of visitors and groups of visitors, including aspects like emotion, mood and personality, and be aware of the context (e.g. weather, security). TheRoute is developed according the Ambient Intelligence perspective. At this point of the project execution we have already developed the full system architecture as a System of Systems approach according to an Ambient Intelligence perspective, to allow the best possible performance in the system utilization for the final user. Intelligent route generation uses user preferences for the categories of points of interest, as well as their personality traits.

- **Algorithms for context-awareness route generation**

  Ricardo Pinto, Luís Conceição, and Goreti Marreiros. "Algorithms for context-awareness route generation". In: *International Symposium on Ambient Intelligence*. Springer. 2020, pp. 93–105

  Abstract. This work has as its main goal the investigation and experimentation on automatic generation of routes for tourists and visitors of points of interest, considering the knowledge of the routes, the profile of the visitor and the context awareness of the tour. The context of a trip can be taken through various sources of information, such as the location of the tourist, the time of the visit, the weather conditions, as well as relevant aspects and characteristics of the user's activity and profile. The developments of this work are part of TheRoute project, and its main goal is the development of a route generation module that considers the context of the tourist, the trip and the environmental constraints. In order to solve the proposed problem, two algorithmic solutions were developed. One is an adaption of the A* algorithm with cuts, while the other is based on Ant Colony Optimization, a Swarm Intelligence algorithm. The results from the experiments allowed to conclude that the A* with cuts, oriented to the heuristic for the path with the highest score, is the one that obtains the best conjugation of results for the defined satisfaction metrics.

- **Requirement Specification for a Remote Monitoring System to Support the Management of Vascular Diseases**

  Sara Escadas, Julio Souza, Ana Vieira, Luís Conceição, Sérgio Sampaio, and Alberto Freitas. "Requirement Specification for a Remote Monitoring System to Support the Management of Vascular Diseases". In: *World Conference on Information Systems and Technologies*. Springer. 2022, pp. 699–710

  Abstract. The development of smart wearable systems for healthcare monitoring has driven extensive efforts from academia and industry in recent years. The present work sought to identify

requirements which are essential for designing a smart wearable remote monitoring system to support the management of selected vascular diseases. To this aim, we reviewed the current research and developments on smart wearable systems for monitoring patients with intermittent claudication (IC), venous ulcers and diabetic foot ulcers (DFU), and a set of functional and technical requirements were extracted and described with the help of clinicians, after studying related projects, clinical scenarios, and healthcare needs. Regarding IC, the requirements focused on measuring the walking capacity and determining the moment at which ambulation cannot continue due to pain. For venous ulcers, sensors will be used to monitor the time spent in a given body position and the pressure applied by compression products. Plantar pressure distribution while standing and temperature will be the main monitored variables for DFU, measured with force and temperature sensors embedded in in-shoe insoles. Coaching services will inform patients and clinicians about the patients' health status, behaviors to adopt and other insights for decision-making support. These users will interact and receive feedback from the system through mobile and web applications. The main innovation aspect of the proposed system consists in a set of intelligent services that allow the smart coaching of patients and healthcare professionals, promoting healthy behaviors and increasing the involvement in treatments, addressing current gaps and needs.

- **Defining an architecture for a remote monitoring platform to support the self-management of vascular diseases**

  Ana Vieira, João Carneiro, Luís Conceição, Constantino Martins, Julio Souza, Alberto Freitas, and Goreti Marreiros. "Defining an architecture for a remote monitoring platform to support the self-management of vascular diseases". In: *Practical Applications of Agents and Multi-Agent Systems*. Springer. 2021, pp. 165–175

  Abstract. The aging of the worldwide population has led to a growing prevalence of vascular diseases, negatively impacting national healthcare systems and patients. The application of information technologies in the health sector has the potential to allow the remote monitoring of patients and the personalization of healthcare. This work proposes the architecture for a remote monitoring platform that supports the self-management of vascular diseases in the context of the Inno4Health project, with the goal of stimulating the innovation in the continuous monitoring of the patients' health. The platform aims to support health professionals and patients with vascular diseases through the continuous monitoring of their health condition and presentation of monitoring reports. Self-management tools will be provided to patients through the presentation of personalized recommendations adapted to their current health condition. By doing so, we believe it will be possible to improve the patients' self-management of their disease as well as decrease the risk of disease-related complications.

- **Implementation of a FHIR Specification for the Interoperability of a Remote Monitoring Platform of Patients with Vascular Diseases**
  Ana Vieira, Luís Conceição, Luiz Faria, Paulo Novais, and Goreti Marreiros. "Implementation of a FHIR Specification for the Interoperability of a Remote Monitoring Platform of Patients with Vascular Diseases". In: *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer. 2022, pp. 246–257
  Abstract. To address the current burdens in the healthcare of patients with vascular diseases, the Portuguese consortium of the Inno4Health project is currently developing a remote monitoring platform that aims to support the self-management of patients with vascular diseases. With the continuous remote monitoring of patients, it is intended that the platform supports both patients and health professionals. Patients will be supported through the presentation of personalized recommendations of activities to perform and health professionals will be supported in the clinical decision making through the presentation of meaningful insights. As the platform is composed by several components, including monitoring sensors, data processing modules and user applications, theressss is a need for the standardization of the data used. This work presents the implementation of a Fast Healthcare Interoperability Resources (FHIR) specification in the remote monitoring platform to ensure the standardization of the data that will be collected and exchanged between the several components of the platform. With this work, it will be possible to easily and securely exchange data, as well as integrating new monitoring sensors and user applications in the platform.

3.2.3 Collaboration and Participation in Scientific Reviews and Events

During this Ph.D., there was the opportunity to participate in and contribute to the organization of national and international scientific events. These opportunities allowed me to live new experiences and learn from other researchers from different locations with different backgrounds and cultures.
Program Committee Participation:

- DeRePAI@PAAMS'23 and DeRePAI@PAAMS'21 - International Conference on Practical Applications of Agents and Multi-Agent System. Workshop on Decision Support, Recommendation, a Workshop on decision support, recommendation, and persuasion in artificial intelligence (DeRePAI 2023). This workshop explores the links between decision support, recommendation, and persuasion to discuss strategies to facilitate the decision/choice process by individuals and groups. It aims to be a discussion forum on the latest trends and ongoing challenges in applying artificial intelligence technologies in this area.

- DiTTEt'22 and DiTTEt'21 - Disruptive Technologies Tech Ethics and Artificial Intelligence. DiTTEt provides a forum to present and discuss the latest scientific and technical advances and their implications in the field of ethics. Also, provide a forum for experts to present their latest research in disruptive technologies, promoting knowledge transfer. It provides a unique opportunity to bring together experts in different fields, academics, and professionals to exchange their experience in the development and deployment of disruptive technologies, artificial intelligence, and ethical problems.

- DeLA@PAAMS'22 - International Conference on Practical Applications of Agents and Multi-Agent System. Workshop on Deep Learning Applications (DeLA). The objective of the workshop on Deep Learning Applications is to give an opportunity for researchers to provide further insight into the problems solved at this stage, the advantages and disadvantages of the various approaches used, lessons learned, and meaningful contributions to enhance applications based on deep learning.

- DICTI'21 and DICTI'19 - Disruptive Information and Communication Technologies for Innovation and digital transformation The workshop on Disruptive Information and Communication Technologies for Innovation and Digital transformation, organized under the scope of the DISRUPTIVE project, aims to discuss problems, challenges, and benefits of using disruptive digital technologies, namely the Internet of Things, Big data, cloud computing, multi-agent systems, machine learning, virtual and augmented reality, and collaborative robotics, to support the on-going digital transformation in society.

- SEI'21, SEI'20 and SEI'19 - Simpósio de Engenharia Informática. It intends to be a space for teachers, researchers and students to share, discuss and reflect on research, development, and practices in the field of Informatics Engineering.

- ISAmI'19 and ISAmI '18 - International Symposium on Ambient Intelligence. ISAmI is the International Symposium on Ambient Intelligence, aiming to bring together researchers from various disciplines that constitute the scientific field of Ambient Intelligence to present and discuss the latest results, new ideas, projects, and lessons learned.

3.2.4 Lecturing

During the doctoral program, the doctoral candidate was invited to be a guest assistant professor at the School of Engineering from the Polytechnic of Porto. The list of lectured courses was the following:

- Social Aspects of Artificial Intelligence (ASPSOCIA) - 2021/2022, 2022/2023: Practical classes;

- Natural Language and Conversational Systems (LNSCIA) - 2020/2021, 2021/2022, 2022/2023: Practical classes;

- Knowledge-Based Systems (SIBAC) - 2020/2021: Practical classes;

- Medical Informatics (INFME) - 2018/2019,2019/2020: Practical classes;

- Enterprise Information Systems (SINFE) - 2018/2019, 2019/2020: Theoretical and Practical classes.

It is important to notice that the two first courses are in the scope of the Master of Engineering and Artificial Intelligence, and are directly connected with the topic of this Ph.D. In the course on Social Aspects of Artificial Intelligence, topics like argumentation, Group Decision-Making, and affective computing are addressed. The course on Natural Language and Conversational Systems addresses topics like text classification, text preprocessing, feature extraction from text, deep learning for text classification; word and sentence embeddings; and conversational systems.

3.2.5 Supervision of Students

During the doctoral program, the doctoral candidate supervised and cosupervised (along with Professor Maria Goreti Carvalho Marreiros), master and undergraduate students that developed work in the context of this thesis at the Polytechnic Institute of Porto. The list of students and their respective dissertation titles was the following:

- Nair Gomes D'Araújo (2022), Master Biomedical Engineering, Dissertation title: "Development of a Nutritional and Physical Activity Plan Recommendation System for Patients with Type 2 diabetes.";
- Vasco Veiga Martins Rodrigues (2022), Master in Computer Engineering, Dissertation title: "Application of clustering techniques in group decision-making context";
- João Ferreira Trindade Mendes Godinho (2020), Master in Computer Engineering and Medical Instrumentation, Dissertation title: "Mobile Application for register of nutrition and physical activity for patients with type 2 diabetes mellitus.";
- Sara Catarina Mendes Batista (2020), Master in Computer Engineering and Medical Instrumentation, Dissertation title: "Nutritional Recommender System for Patients with Type 2 Diabetes Mellitus.";
- João Tiago da Mota Moreira (2020), Master in Computer Engineering, Dissertation title: "Deteção de emoções em argumentos trocados no context da tomada de decisão em grupo";
- Ricardo Emanuel Capelas Pinto (2019), Master in Computer Engineering, Dissertation title: "Algoritmos para geração de rotas com base na análise do contexto";
- João Miguel Duarte Couto (2019), Bachelor in Informatics Engineering, Project title: "Plataforma para gestão de recursos humanos em projetos de investigação científica";
- Célio Eduardo Sousa Cerqueira (2019), Bachelor in Informatics Engineering, Project title: Desenvolvimento de um Sistema de apoio à decisão em grupo baseado na web;
- Rodrigo Fernandes da Palma Mesquita (2018), Bachelor in Informatics Engineering, Project title: "Criação automática de grupos de turistas e agregação de preferências considerando a componente afetiva".

3.2.6 Awards and Recognition

During the development of this thesis, the candidate saw his work being awarded as a distinction of the quality of work, having got the distinctions listed below:

- Award "Second Best Application Paper in the International Conference ISAMI'2018"with the work Carlos Ramos, Goreti Marreiros, Constantino Martins, Luiz Faria, Luís Conceição, Joss Santos, Luís Ferreira, Rodrigo Mesquita, and Lucas Schwantes Lima. "A context-awareness approach to tourism and heritage routes generation". In: *International Symposium on Ambient Intelligence*. Springer. 2018, pp. 10–23

- Master thesis supervisor of "Best Student Award of the Master in Informatics"with Ricardo Emanuel Capelas Pinto (2019), with the work "Algoritmos para geração de rotas com base na análise do contexto".

3.3 Final Remarks and Future Work Considerations

GDSS and the web-based GDSS have evolved since their appearance in the '80s, and the benefits for the group decision-making processes in their utilization are identified in the literature. Despite that, there is still some space for improvements to attract and respond to the necessities identified by the decision-makers.

The contributions of this Ph.D. thesis are an increment to the state of the art of the web-based Group Decision Support Systems namely in the application of argument clustering techniques. The developed models are a small but important contribution to the development of web-based GDSS.

Models to perform automatic classification (InFavour/Against) were applied to argumentative sentences exchanged in GDSS's by decision-makers, to dynamically organize the conversation based on the arguments used, allowing intelligent reports for GDSS [19, 21] to become even more helpful since they can dynamically organize the conversations between decision-makers by grouping them based on, for instance, the alternatives, the criteria, and comparisons, among others, that are being discussed. Also, a web-based GDSS prototype based on a multi-agent system was designed. This is an important step to increase the acceptance of these systems in organizations, bringing the advantages of face-to-face meetings to the online scenario.

As future work, it is important to mention that the main lines of the work developed during this doctoral program will have applicability and continuity in research projects like PRR: Agenda to Accelerate and Transform Tourism, or smarTravel.

PRR: Agenda to Accelerate and Transform Tourism The consortium is composed of 44 entities, jointly promoted by the Confederation of Tourism of Portugal (CTP), by Turismo de Portugal, and by NEST – Center for Tourism Innovation, the "Agenda Accelerate and Transform Tourism" foresees an investment of 151 million euros in a set of transformative innovation projects with particular emphasis on digitalization and climate transition. ISEP will focus on developing technologies focused on the tourist, namely

proposing recommendations to groups of tourists that will be generated with clustering techniques. The investments in this Agenda, are focused on the tourist, to respond to the competitive situation between tourist destinations, which implies a new standard of offering throughout the tourist journey, to reach a more excellent level of efficiency and value proposal suitable for the new standards of consumption and sustainability. This project was approved in the scope of Plano de Recuperação e Resiliência (PRR), in the call 02/C05-i01/2022.

The smarTravel - Smart Travel Digital Ecosystem - project consists of the development of a digital ecosystem integrated into the concept of smart cities with a view to improving all phases involving the tourist experience, based on the application of technologies such as Artificial Intelligence, continuous extraction of information, characterization of behaviors, new travel trends, seasonal choices, as well as choices directly linked to the weather and trip optimization. Thus, the intention is to develop a system capable of assisting user-tourists in the stages of idealization, planning, booking, transport, experience, and sharing, allowing, in each of these stages, to analyze user profiles, their preferences in terms of trip and pre-existing conditions to create personalized travel recommendations (useful and tailored to each specific user), making the whole process more pleasant and efficient, considering, among others, its duration, weather conditions, cost and customer expectations. user. This project is funded by Compete 2020 with the reference: POCI-01-0247-FEDER-179946.

Bibliography

[1] Leila Amgoud. “Postulates for logic-based argumentation systems”. In: *International Journal of Approximate Reasoning* 55.9 (2014), pp. 2028–2048. issn: 0888-613X.

[2] Leila Amgoud and Henri Prade. “Explaining Qualitative Decision under Uncertainty by Argumentation”. In: *Proceedings of the National Conference on Artificial Intelligence*. Vol. 21. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006, p. 219.

[3] Leila Amgoud and Srdjan Vesic. “A formal analysis of the outcomes of argumentation-based negotiations”. In: *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*. 2011, pp. 1237–1238.

[4] Katie Atkinson and Trevor Bench-Capon. “Practical reasoning as presumptive argumentation using action based alternating transition systems”. In: *Artificial Intelligence* 171.10-15 (2007), pp. 855–874.

[5] Lawrence Birnbaum, Margot Flowers, and Rod McGuire. “Towards an AI model of argumentation”. In: *Proceedings of the First AAAI Conference on Artificial Intelligence*. 1980, pp. 313–315.

[6] Tom Bosc, Elena Cabrio, and Serena Villata. “Tweeties squabbling: Positive and negative results in applying argument mining on social media.” In: *COMMA* 2016 (2016), pp. 21–32.

[7] Tiago Cardoso, Vasco Rodrigues, Luís Conceição, João Carneiro, Goreti Marreiros, and Paulo Novais. “Aspect Based Sentiment Analysis Annotation Methodology for Group Decision Making Problems: An Insight on the Baseball Domain”. In: *World Conference on Information Systems and Technologies*. Springer. 2022, pp. 25–36.

[8] João Carneiro, Patrícia Alves, Goreti Marreiros, and Paulo Novais. “Group decision support systems for current times: Overcoming the challenges of dispersed group decision-making”. In: *Neurocomputing* 423 (2021), pp. 735–746.

[9] João Carneiro, Rui Andrade, Patrícia Alves, Luís Conceição, Paulo Novais, and Goreti Marreiros. “A consensus-based group decision support system using a multi-agent MicroServices approach”. In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. 2020, pp. 2098–2100.

[10] João Carneiro, Luís Conceição, Diogo Martinho, Goreti Marreiros, and Paulo Novais. "Including cognitive aspects in multiple criteria decision analysis". In: *Annals of Operations Research* 265.2 (2018), pp. 269–291.

[11] João Carneiro, Diogo Martinho, Patrícia Alves, Luís Conceição, Goreti Marreiros, and Paulo Novais. "A multiple criteria decision analysis framework for dispersed group decision-making contexts". In: *Applied Sciences* 10.13 (2020), p. 4614.

[12] João Carneiro, Diogo Martinho, Goreti Marreiros, Amparo Jimenez, and Paulo Novais. "Dynamic argumentation in UbiGDSS". In: *Knowledge and Information Systems* 55.3 (2018), pp. 633–669. issn: 0219-1377.

[13] João Carneiro, Diogo Martinho, Goreti Marreiros, and Paulo Novais. "A General Template to Configure Multi-Criteria Problems in Ubiquitous GDSS". In: *International Journal of Software Engineering and Its Applications* 9 (2015), pp. 193–206. doi: 10.14257/astl.205.97.17.

[14] João Carneiro, Pedro Saraiva, Luís Conceição, Ricardo Santos, Goreti Marreiros, and Paulo Novais. "Predicting satisfaction: Perceived decision quality by decision-makers in Web-based group decision support systems". In: *Neurocomputing* (2019). issn: 0925-2312. doi: doi.org/10.1016/j.neucom.2018.05.126.

[15] Lucas Carstens and Francesca Toni. "Towards relation based argumentation mining". In: *Proceedings of the 2nd Workshop on Argumentation Mining*. 2015, pp. 29–34.

[16] Lucas Carstens and Francesca Toni. "Using argumentation to improve classification in natural language problems". In: *ACM Transactions on Internet Technology (TOIT)* 17.3 (2017), pp. 1–23.

[17] Chen-Tung Chen. "Extensions of the TOPSIS for group decision-making under fuzzy environment". In: *Fuzzy sets and systems* 114.1 (2000), pp. 1–9. issn: 0165-0114.

[18] Oana Cocarascu and Francesca Toni. "Identifying attack and support argumentative relations using deep learning". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2017, pp. 1374–1379.

[19] Luís Conceição, João Carneiro, Goreti Marreiros, and Paulo Novais. "A SOA Web-Based Group Decision Support System Considering Affective Aspects". In: *EPIA Conference on Artificial Intelligence*. Springer. 2017, pp. 65–74.

[20] Luís Conceição, João Carneiro, Goreti Marreiros, and Paulo Novais. "Applying Machine Learning Classifiers in Argumentation Context". In: *International Symposium on Distributed Computing and Artificial Intelligence*. Springer. 2020, pp. 314–320.

[21] Luís Conceição, João Carneiro, Diogo Martinho, Goreti Marreiros, and Paulo Novais. "Generation of intelligent reports for ubiquitous group decision support systems". In: *2016 Global Information Infrastructure and Networking Symposium (GIIS)*. IEEE. 2016, pp. 1–6.

[22] Luís Conceição, Diogo Martinho, Rui Andrade, João Carneiro, Constantino Martins, Goreti Marreiros, and Paulo Novais. “A web-based group decision support system for multicriteria problems”. In: *Concurrency and Computation: Practice and Experience* 33.2 (2021), e5298.

[23] Luís Conceição, Vasco Rodrigues, Jorge Meira, Goreti Marreiros, and Paulo Novais. “Supporting Argumentation Dialogues in Group Decision Support Systems: An Approach Based on Dynamic Clustering”. In: *Applied Sciences* 12.21 (2022), p. 10893.

[24] Luís Manuel da Silva Conceição. “Sistema de Apoio à tomada de Decisão em Grupo baseado na Web que considera aspetos cognitivos”. 2017.

[25] Artur S D’Avila Garcez, Dov M Gabbay, and Luis C Lamb. “Value-based argumentation frameworks as neural-symbolic learning systems”. In: *Journal of Logic and Computation* 15.6 (2005), pp. 1041–1058. issn: 1465-363X.

[26] Ido Dagan, Oren Glickman, and Bernardo Magnini. “The pascal recognising textual entailment challenge”. In: *Machine learning challenges workshop*. Springer. 2006, pp. 177–190.

[27] Gerardine DeSanctis and Brent Gallupe. “Group decision support systems: a new frontier”. In: *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 16.2 (1984), pp. 3–10.

[28] Gerardine DeSanctis and R Brent Gallupe. “A foundation for the study of group decision support systems”. In: *Management science* 33.5 (1987), pp. 589–609.

[29] Ru-Xi Ding, Iván Palomares, Xueqing Wang, Guo-Rui Yang, Bingsheng Liu, Yucheng Dong, Enrique Herrera-Viedma, and Francisco Herrera. “Large-Scale decision-making: Characterization, taxonomy, challenges and future directions from an Artificial Intelligence and applications perspective”. In: *Information Fusion* 59 (2020), pp. 84–102.

[30] Ru-Xi Ding, Xueqing Wang, Kun Shang, and Francisco Herrera. “Social network analysis-based conflict relationship investigation and conflict degree-based consensus reaching process for large scale decision making using sparse representation”. In: *Information Fusion* 50 (2019), pp. 251–272.

[31] Yucheng Dong, Sihai Zhao, Hengjie Zhang, Francisco Chiclana, and Enrique Herrera-Viedma. “A self-management mechanism for noncooperative behaviors in large-scale group consensus reaching processes”. In: *IEEE Transactions on Fuzzy Systems* 26.6 (2018), pp. 3276–3288.

[32] Phan Minh Dung. “On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games”. In: *Artificial intelligence* 77.2 (1995), pp. 321–357. issn: 0004-3702.

[33] Sara Escadas, Julio Souza, Ana Vieira, Luís Conceição, Sérgio Sampaio, and Alberto Freitas. “Requirement Specification for a Remote Monitoring System to Support the Management of Vascular Diseases”. In: *World Conference on Information Systems and Technologies*. Springer. 2022, pp. 699–710.

[34] Andrea Esuli and Fabrizio Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining". In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. 2006.

[35] Xiuyi Fan and Francesca Toni. "Decision making with assumption-based argumentation". In: *International Workshop on Theorie and Applications of Formal Argumentation*. Springer, 2013, pp. 127–142.

[36] Margot Flowers, Rod McGuire, and Lawrence Birnbaum. "Adversary arguments and the logic of personal attacks". In: *Strategies for natural language processing*. Psychology Press, 2014, pp. 297–316.

[37] Diego Garcia-Zamora, Álvaro Labella, Weiping Ding, Rosa M Rodríguez, and Luis Martínez. "Large-Scale Group Decision Making: A Systematic Review and a Critical Analysis". In: *IEEE/CAA Journal of Automatica Sinica* 9.6 (2022), pp. 949–966.

[38] Seyed Morsal Ghavami, Mohammad Taleai, and Theo Arentze. "An intelligent web-based spatial group decision support system to investigate the role of the opponents' modeling in urban land use planning". In: *Land Use Policy* 120 (2022), p. 106256.

[39] Nabila Hadidi, Yannis Dimopoulos, and Pavlos Moraitis. "Argumentative alternating offers". In: *International Workshop on Argumentation in Multi-Agent Systems*. Springer. 2011, pp. 105–122.

[40] Francisco Herrera and Enrique Herrera-Viedma. "Linguistic decision analysis: steps for solving decision problems under linguistic information". In: *Fuzzy Sets and systems* 115.1 (2000), pp. 67–82.

[41] Francisco Herrera, L Martınez, and Pedro J Sánchez. "Managing non-homogeneous information in group decision making". In: *European Journal of Operational Research* 166.1 (2005), pp. 115–132. issn: 0377-2217.

[42] Jos van Hillegersberg and Sebastiaan Koenen. "Adoption of web-based group decision support systems: Conditions for growth". In: *Procedia technology* 16 (2014), pp. 675–683. issn: 2212-0173.

[43] Jos van Hillegersberg and Sebastiaan Koenen. "Adoption of web-based group decision support systems: experiences from the field and future developments". In: *International Journal of Information Systems and Project Management* 4 (2016), pp. 49–64.

[44] George P. Huber. "Issues in the Design of Group Decision Support Systems". In: *MIS Quarterly: Management Information Systems* 8 (1984), pp. 195–204.

[45] Sam Kaner. *Facilitator's guide to participatory decision-making*. John Wiley & Sons, 2014. isbn: 1118404955.

[46] Sarit Kraus, Katia Sycara, and Amir Evenchik. "Reaching agreements through argumentation: a logical model and implementation". In: *Artificial Intelligence* 104.1 (1998), pp. 1–69. issn: 0004-3702.

[47] John Lawrence and Chris Reed. "Argument mining: A survey". In: *Computational Linguistics* 45.4 (2020), pp. 765–818.

[48] Kurt Lewin. "Action Research and Minority Problems". In: *Journal of Social Issues* 2.4 (1946), pp. 34–46. doi: https://doi.org/10.1111/j.1540-4560.1946.tb02295.x. eprint: https://spssi.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-4560.1946.tb02295.x. url: https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4560.1946.tb02295.x.

[49] Cong-Cong Li, Yucheng Dong, and Francisco Herrera. "A consensus model for large-scale linguistic group decision making with a feedback recommendation based on clustered personalized individual semantics and opposing consensus groups". In: *IEEE Transactions on Fuzzy Systems* 27.2 (2018), pp. 221–233.

[50] Bingsheng Liu, Yinghua Shen, Xiaohong Chen, Yuan Chen, and Xueqing Wang. "A partial binary tree DEA-DA cyclic classification model for decision makers in complex multi-attribute large-group interval-valued intuitionistic fuzzy decision-making problems". In: *Information Fusion* 18 (2014), pp. 119–130.

[51] Bingsheng Liu, Qi Zhou, Ru-Xi Ding, Wei Ni, and Francisco Herrera. "Defective alternatives detection-based multi-attribute intuitionistic fuzzy large-scale decision making model". In: *Knowledge-Based Systems* 186 (2019), p. 104962.

[52] João M. Lourenço. *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. 2021. url: https://github.com/joaomlourenco/novathesis/raw/master/template.pdf.

[53] Fred Luthans, Brett C Luthans, and Kyle W Luthans. *Organizational Behavior: An evidencebased approach*. IAP, 2015. isbn: 1681231212.

[54] Omar Marey, Jamal Bentahar, Ehsan Khosrowshahi Asl, Mohamed Mbarki, and Rachida Dssouli. "Agents' uncertainty in argumentation-based negotiation: classification and implementation". In: *Procedia Computer Science* 32 (2014), pp. 61–68. issn: 1877-0509.

[55] Goreti Marreiros, Ricardo Santos, Paulo Novais, José Machado, Carlos Ramos, José Neves, and José Bula-Cruz. "Argumentation-based decision making in ambient intelligence environments". In: *Portuguese Conference on Artificial Intelligence*. Springer. 2007, pp. 309–322.

[56] Goreti Marreiros, Ricardo Santos, Carlos Ramos, and José Neves. "Context-aware emotion-based model for group decision making". In: *IEEE Intelligent Systems* 25.2 (2010), pp. 31–39.

[57] Maria Goreti Carvalho Marreiros. “Agentes de apoio à argumentação e decisão em grupo”. PhD thesis. 2007.

[58] George A Miller. “WordNet: a lexical database for English”. In: *Communications of the ACM* 38.11 (1995), pp. 39–41.

[59] Miguel Miranda, António Abelha, Manuel Santos, José Machado, and José Neves. “A group decision support system for staging of cancer”. In: *Electronic Healthcare*. Springer, 2008, pp. 114–121. isbn: 3642004121.

[60] Juan Antonio Morente-Molinera, Robin Wikström, Enrique Herrera-Viedma, and Christer Carlsson. “A linguistic mobile decision support system based on fuzzy ontology to facilitate knowledge mobilization”. In: *Decision Support Systems* 81 (2016), pp. 66–75. issn: 0167-9236.

[61] Gonzalo Navarro. “A guided tour to approximate string matching”. In: *ACM computing surveys (CSUR)* 33.1 (2001), pp. 31–88.

[62] Iván Palomares, Luis Martinez, and Francisco Herrera. “A consensus model to detect and manage noncooperative behaviors in large-scale group decision making”. In: *IEEE Transactions on Fuzzy Systems* 22.3 (2013), pp. 516–530.

[63] Iván Palomares and Luis Martínez. “A semisupervised multiagent system model to support consensus-reaching processes”. In: *IEEE Transactions on Fuzzy Systems* 22.4 (2014), pp. 762–777. issn: 1063-6706.

[64] Iván Palomares, Francisco J Quesada, and Luis Martínez. “An approach based on computing with words to manage experts behavior in consensus reaching processes with large groups”. In: *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE. 2014, pp. 476–483.

[65] Ricardo Pinto, Luís Conceição, and Goreti Marreiros. “Algorithms for context-awareness route generation”. In: *International Symposium on Ambient Intelligence*. Springer. 2020, pp. 93–105.

[66] Henry Prakken. “An abstract framework for argumentation with structured arguments”. In: *Argument and Computation* 1.2 (2010), pp. 93–124. issn: 1946-2166.

[67] Francisco J Quesada, Ivan Palomares, and Luis Martinez. “Managing experts behavior in large-scale consensus reaching processes with uninorm aggregation operators”. In: *Applied Soft Computing* 35 (2015), pp. 873–887.

[68] Francisco José Quesada, Iván Palomares, and Luis Martínez. “Using computing with words for managing non-cooperative behaviors in large scale group decision making”. In: *Granular Computing and Decision-Making*. Springer, 2015, pp. 97–121.

[69] Iyad Rahwan and Leila Amgoud. “An argumentation-based approach for practical reasoning”. In: *International Workshop on Argumentation in Multi-Agent Systems*. Springer. 2006, pp. 74–90.

[70] Iyad Rahwan, Sarvapali D Ramchurn, Nicholas R Jennings, Peter Mcburney, Simon Parsons, and Liz Sonenberg. "Argumentation-based negotiation". In: *The Knowledge Engineering Review* 18.04 (2003), pp. 343–375. issn: 1469-8005.

[71] Carlos Ramos, Goreti Marreiros, Constantino Martins, Luiz Faria, Luís Conceição, Joss Santos, Luís Ferreira, Rodrigo Mesquita, and Lucas Schwantes Lima. "A context-awareness approach to tourism and heritage routes generation". In: *International Symposium on Ambient Intelligence*. Springer. 2018, pp. 10–23.

[72] Ariel Rosenfeld and Sarit Kraus. "Providing arguments in discussions on the basis of the prediction of human argumentative behavior". In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6.4 (2016), p. 30. issn: 2160-6455.

[73] Ariel Rosenfeld and Sarit Kraus. "Strategical argumentative agent for human persuasion". In: *Proceedings of the Twenty-second European Conference on Artificial Intelligence*. IOS Press, 2016, pp. 320–328. isbn: 1614996717.

[74] Ariel Rosenfeld and Sarit Kraus. "Strategical argumentative agent for human persuasion: A preliminary report". In: *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing*. 2016.

[75] Thomas L Saaty. "What is the analytic hierarchy process?" In: *Mathematical models for decision support*. Springer, 1988, pp. 109–121.

[76] Thomas L. Saaty. "Decision Making with the Analytic Hierarchy Process". In: *International Journal of Services Sciences* 1 (2008), pp. 83–98.

[77] Carles Sierra, Nick R Jennings, Pablo Noriega, and Simon Parsons. "A framework for argumentation-based negotiation". In: *International Workshop on Agent Theories, Architectures, and Languages*. Springer, 1997, pp. 177–192.

[78] Ashraf B El-Sisi and Hamdy M Mousa. "Argumentation based negotiation in multiagent system". In: *2012 Seventh International Conference on Computer Engineering & Systems (ICCES)*. IEEE. 2012, pp. 261–266.

[79] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. "Recursive deep models for semantic compositionality over a sentiment treebank". In: *Proceedings of the 2013 conference on empirical methods in natural language processing*. 2013, pp. 1631–1642.

[80] Ming Tang and Huchang Liao. "From conventional group decision making to large-scale group decision making: What are the challenges and how to meet them in big data era? A state-of-the-art survey". In: *Omega* 100 (2021), p. 102141.

[81] Ming Tang, Xiaoyang Zhou, Huchang Liao, Jiuping Xu, Hamido Fujita, and Francisco Herrera. "Ordinal consensus measure with objective threshold for heterogeneous large-scale group decision making". In: *Knowledge-Based Systems* 180 (2019), pp. 62–74.

[82] Madjid Tavana, Majid Behzadian, Mohsen Pirdashti, and Hasan Pirdashti. "A PROMETHEE-GDSS for oil and gas pipeline planning in the Caspian Sea basin". In: *Energy Economics* 36 (2013), pp. 716–728. issn: 0140-9883.

[83] Alice Toniolo, Timothy J Norman, and Katia P Sycara. "An empirical study of argumentation schemes in deliberative dialogue". In: *ECAI 2012* (2012).

[84] Frans H Van Eemeren, Rob Grootendorst, Ralph H Johnson, Christian Plantin, and Charles A Willard. *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Routledge, 2013.

[85] Ana Vieira, João Carneiro, Luís Conceição, Constantino Martins, Julio Souza, Alberto Freitas, and Goreti Marreiros. "Defining an architecture for a remote monitoring platform to support the self-management of vascular diseases". In: *Practical Applications of Agents and Multi-Agent Systems*. Springer. 2021, pp. 165–175.

[86] Ana Vieira, Luís Conceição, Luiz Faria, Paulo Novais, and Goreti Marreiros. "Implementation of a FHIR Specification for the Interoperability of a Remote Monitoring Platform of Patients with Vascular Diseases". In: *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer. 2022, pp. 246–257.

[87] Douglas Walton and Erik CW Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press, 1995. isbn: 079142586X.

[88] Pei Wang, Xuanhua Xu, Shuai Huang, and Chenguang Cai. "A linguistic large group decision making method based on the cloud model". In: *IEEE Transactions on Fuzzy Systems* 26.6 (2018), pp. 3314–3326.

[89] Tong Wu, Xinwang Liu, and Fang Liu. "An interval type-2 fuzzy TOPSIS model for large scale group decision making problems with social network information". In: *Information Sciences* 432 (2018), pp. 392–410.

[90] Tong Wu, Xinwang Liu, and Jindong Qin. "A linguistic solution for double large-scale group decision-making in E-commerce". In: *Computers & Industrial Engineering* 116 (2018), pp. 97–112.

[91] Xuan-hua Xu, Zhi-jiao Du, and Xiao-hong Chen. "Consensus model for multi-criteria large-group emergency decision making considering non-cooperative behaviors and minority opinions". In: *Decision Support Systems* 79 (2015), pp. 150–160.

[92] Xuan-hua Xu, Zhi-jiao Du, Xiao-hong Chen, and Chen-guang Cai. "Confidence consensus-based model for large-scale group decision making: A novel approach to managing non-cooperative behaviors". In: *Information Sciences* 477 (2019), pp. 410–427.

[93] Yejun Xu, Xiaowei Wen, and Wancheng Zhang. "A two-stage consensus method for large-scale multi-attribute group decision making with an application to earthquake shelter selection". In: *Computers & Industrial Engineering* 116 (2018), pp. 113–129.

[94] Morteza Yazdani, Pascale Zarate, Adama Coulibaly, and Edmundas Kazimieras Zavadskas. "A group decision making support system in logistics and supply chain management". In: *Expert Systems with Applications* 88 (2017), pp. 376–392. issn: 0957-4174.

[95] Cristina Zuheros, Eugenio Martínez-Cámara, Enrique Herrera-Viedma, and Francisco Herrera. "Sentiment analysis based multi-person multi-criteria decision making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews". In: *Information Fusion* 68 (2021), pp. 22–36.