A Group Decision-Making Method Based on the Experts' Behaviour During the Debate

José Ramón Trillo, Enrique Herrera-Viedma, *Fellow, IEEE*, Juan Antonio Morente-Molinera, and Francisco Javier Cabrerizo

***Abstract*—Debate is a process consisting in arriving at a reasoned opinion on a proportion in which individuals must be truly capable of defending their own judgments. It has been used within group decision-making (GDM) problems to help experts make better decisions. However, whether experts engage in a vigorous debate, it can result in the use of aggressive language that may diminish consensus, which is the major objective of GDM. To avoid it, we present a novel method for GDM problems that can identify aggressive comments during the debate by incorporating a classifier based on sentiment analysis techniques. According to the information extracted during the debate, two procedures are developed to assign weights to the experts, which are used to introduce two new consensus measures and to make the final decision. Unlike the existing GDM methods, this new one can take advantage of the information extracted during the debate (i.e., experts' behaviour) throughout the decision process, making it in rapport with real-world GDM processes.**

***Index Terms*—Group decision-making, consensus, sentiment analysis.**

I. Introduction

Making decisions is a common task, from the professional to the personal level, that involves emotions [1]. Due to it, individuals sometimes make wrong decisions because of their mood or the mood of others [2], [3]. This leads to not choosing the best possible course of action when individuals' feelings cloud their judgments or they suffer regret or discomfort [4], which is a common issue when individuals face group decision-making (GDM) processes [5].

In GDM processes, a group of individuals, usually experts in the problem under discussion, are generally involved in a debate in which they give their assessments on a set of possible alternatives or courses of action and provide arguments supporting them [6]. The objective is to attain a consensual solution that takes advantage of all individuals' ideas [7]. It is due to individuals are more likely to implement decisions they accept, and consensus makes acceptance more likely [8].

The existing approaches supporting GDM processes assist experts in choosing the best solution within a set of possible alternatives or courses of action. Most of these approaches focus on helping experts reach the highest consensus possible before making the final decision [9]–[11]. To do it, they are usually based on the assessments provided by the experts, but not considering the arguments conveyed by them to express such assessments. However, when an expert expresses an assessment, for instance, "this wine brand is much better than that other," it is not only important the given assessment but also the arguments provided to support such assessment [12].

Experts provide arguments to justify their assessments, and it can make known information that is relevant to the GDM process. Firstly, the style of the speaker may be the tiebreaker if experts are equally matched in refutation and argumentation [8]. Second, aggressive language has been shown to negatively affect the reliability of the speaker using it and the credibility of his or her information [13], and consensus demands a high level of trust among the members of the group. Third, aggressive language can also lead to the deterioration of the debate, making experts get defensive and emotional. As a result, they could provide unsuitable assessments [14]. In summary, if the experts' behaviour, in the sense of the language used by the experts to argue their assessments, during the debate is not analysed, information, which may be relevant to the decision-making process, will be lost.

On the one hand, several approaches have been proposed to analyse the debate carried out between the experts before providing their assessments [15]–[19]. However, these approaches do not consider the experts' behaviour during the debate. On the other hand, several approaches have been proposed to take into account the experts' behaviour during the decision-making process [20]–[23], but from a different point of view. Concretely, these approaches deal with non-cooperative behaviours in the sense that experts with such behaviour try to obstruct the improvement of consensus by providing assessments very different from the ones expressed by the rest of the group members. To handle such behaviours, these approaches generally penalize experts with non-cooperative behaviours during the decision-making process. Therefore, the approaches analyzing the debate carried out between the experts in GDM processes should consider both the assessments provided by the experts during the debate and the language used to express their arguments because this information could be of interest to the rest of the decision-making process.

The objective of this study is to develop a new method for GDM processes considering the experts' behaviour in the debate throughout the decision-making process. Firstly, it incorporates a classifier based on sentiment analysis techniques [24] that can categorize the language used by the experts during the debate as aggressive or non-aggressive. Several data preprocessing techniques are applied to extract many

This work was supported by the project PID2019-103880RB-I00 funded by MCIN / AEI / 10.13039/501100011033, by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades / Proyecto B-TIC-590-UGR20, and by the Andalusian Government through the project P20_00673.

José Ramón Trillo, Enrique Herrera-Viedma, Juan Antonio Morente-Molinera, and Francisco Javier Cabrerizo are with the Department of Computer Science and Artificial Intelligence, Andalusian Research Institute of Data Science and Computational Intelligence, DaSCI, University of Granada, Granada, 18071, Spain (e-mail: jrtrillo@ugr.es, viedma@decsai.ugr.es, jamoren@decsai.ugr.es, cabrerizo@decsai.ugr.es).

parameters (number of insults, number of words in capital letters, and so on) allowing to obtain a better classification. Secondly, based on the information generated by the classifier, two new procedures are designed to assign weights to the experts, which are used both to define two new consensus measures and to make the final decision. In consequence, this new method improves the existing ones by taking advantage of the information extracted in the debate (i.e., the experts' behaviour) during the whole decision-making process.

The content of this study is divided into five sections. In Section II, we recall the basic concepts and techniques associated with the proposed method. In Section III, we elaborate on the new method for GDM. Strong attention is given to how the experts' behaviour is extracted during the debate and how this information is used in the rest of the decision-making process. Section IV provides a real-world case study that helps understand the performance of the proposed method. Finally, Section V provides some concluding remarks about the advantages and disadvantages of the proposed method along with some further research directions.

II. Background

In this section, we recall the necessary background to understand the GDM method developed in this study. Section II-A is dedicated to sentiment analysis, in particular, the steps needed to build a classifier are described. Section II-B is devoted to introducing the basic concepts concerning GDM problems.

A. Sentiment Analysis

Natural language, which is used by humans for communication, is extraordinarily diverse and complex. It is composed of vague and imprecise concepts that are far from the binary language used by computers, in which communication occurs through millions of ones and zeros, instead of words, to produce logical actions. Therefore, to help computers understand humans in their own language, natural language processing is required [25], [26].

When individuals hold a debate on a topic, they tend to use natural language. Consequently, the use of tools translating this language into one that can be understood by a computer is necessary. These tools can also help extract useful information from the data. For instance, in GDM processes, it is interesting to know how experts feel. This can be done by analyzing the language employed by the experts during the debate to gain insights into their sentiments and emotions, which can reveal the experts' behaviour during the debate. To do this, sentiment analysis, one of the most important fields of natural language processing, can be used [24], [27].

Two distinct approaches using machine learning can be implemented to perform sentiment analysis [28]: a supervised learning approach, which depends on the existence of labelled training data, and an unsupervised learning approach, which uses libraries or tools to classify opinions without an already labelled output. In this study, we assume a supervised learning approach as it is the most basic approach and offers good performance.

Classifiers based on supervised learning for sentiment analysis make use of two data sets. The first one, called training data set, helps the classifier knows which concepts are associated with a feeling and which ones are not. The second one, called test data set, allows us to know whether the classifier detects those feelings appropriately. Based on both data sets, the steps for building a classifier for sentiment analysis are [24]:

- Preprocessing. The first step to be done is the preprocessing step, which is very important in sentiment analysis as it cleans the data sets from any noisy data, thus reducing their complexity to prepare the data for the next steps.
- Training. The second step consists in taking content that is known to belong to a specific class (training data set) and creating a classifier based on that known content employing machine learning algorithms. It is an iterative process whereby the best classifier possible is built.
- Classification. The third step consists in taking the classifier built with such training data set and running it on unknown content to determine class membership for the unknown content (test data set). It is a one-time process designed to run on unknown content.

In recent literature, there exist plenty of articles that use sentiment analysis. For instance, in [29], a method for analysing investor behaviour is presented. For this purpose, it analyses the sentiments through the phrases that investors make during their buying and selling process. In [30], it is presented a hardware framework for optimising text sentiment analysis using a memristor-based long short-term memory system. And, in [31], bidirectional recurring units are used by the process to capture contextual information when learning word representations in sentiment analysis.

B. GDM Processes

A GDM process, otherwise called a collective decision-making problem, characterizes a process in which several individuals (experts) collectively arrive at a choice from some alternatives. It is formally characterized by a finite collection of alternatives, $X = \{x_1, x_2, \ldots, x_n\}$, and a finite group of experts, $E = \{e_1, e_2, \ldots, e_m\}$, where every expert, based on her or his knowledge of the problem under study, provides her or his assessments of the alternatives. The objective is to build a collective ranking of the alternatives from best to worst as a possible solution to the problem [5], [32].

The experts' assessments of the alternatives can represent either the preference degree of one alternative over other one or the degree up to which an alternative is chosen as possible solution to the problem. In any case, a particular representation domain must be selected to model the assessments. Notably, fuzzy set theory has been successfully used in the resolution of decision-making processes because they are cognitive processes in which humans (experts) take part [33]. In fact, the fuzzy set theory and its extensions have been demonstrated to be useful in characterizing decision information pervaded with human uncertainty [34].

To solve a GDM problem, different stages are carried out, which start with the experts holding a debate and providing their assessments of the alternatives and end by deriving a ranking of the alternatives [10]:

- Holding a debate. The experts discuss the different alternatives related to the problem. During this debate, the experts give reasons for their assessments and share their knowledge about the alternatives and the problem to make easier the procedure of afterwards providing their assessments [6].
- Providing assessments. The experts express their assessments of the alternatives employing a certain representation format and a particular representation domain [35].
- Analyzing consensus. The experts must reach enough agreement before making the final decision. It means that we need to compute some consensus measures allowing us to identify the consensus achieved [36]–[38]. The greater the consensus, the more reliable the decision made. On the contrary, if the consensus achieved is not enough, the experts should reconsider their assessments and, then, the above stages must be carried out again [39], [40]. To avoid a cyclical and never-ending process, a limited number of consensus rounds is generally established. When the limit is reached, the ranking of the alternatives must be obtained independently of the consensus achieved.
- Computing a collective assessment. The individual assessments provided by the experts are aggregated to compute a collective assessment summarizing them. To do so, different aggregation functions can be used [41].
- Ranking the alternatives. The information contained in the collective assessment is exploited to rank the alternatives. Different functions can be applied to do it [42], [43]. According to this rank, the best alternative is selected as a solution to the decision problem. This stage is also known as the exploitation stage.

III. A GDM Method Based on Experts' Behaviour in the Debate

In this section, we describe a new method for GDM problems that considers the experts' behaviour during the debate stage throughout the GDM process. This new method makes use of sentiment analysis techniques to identify the experts' feelings through the comments provided by them to support their assessments during the debate, i.e., it can detect the experts' conduct, allowing us to understand how experts feel. It implies that we can know some important information, such as how the decision to be made affects the experts [44], [45].

This new GDM method incorporates a sentiment analysis classifier capable of detecting aggressive language in the experts' debate. It is composed of the following stages (see Fig. 1): (i) detecting aggressive language during the debate, (ii) providing assessments, (iii) computing experts' weights, (iv) analyzing consensus, (v) computing a collective assessment, and (vi) ranking the alternatives. In the following subsections, we elaborate on in detail each of the stages.

A. Detecting Aggressive Language During the Debate

The first stage is devoted to identifying aggressive language in the comments provided by the experts while they debate.

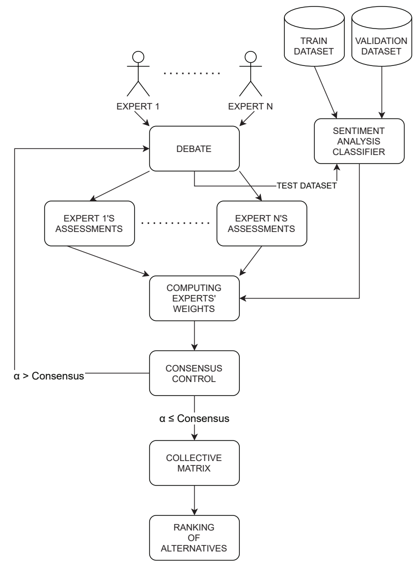

Fig. 1. General outline of the proposed GDM method.

This is done by means of a classifier based on supervised learning for sentiment analysis. As described in Section II-A, to build the classifier, different machine learning algorithms, namely, support vector machine (SVM), Naïve Bayes, k-nearest neighbours (KNN), and so on, can be used [46]. In addition, a training data set is needed to adjust the parameters of the classifier and, once this has been done, a test data set must be used to verify the quality of the classifier.

Once the classifier has been constructed, this stage is carried out as follows. Several debate rounds, $\zeta \in \mathbb{N}$, are performed. In every debate round, first, the comments provided by the experts are preprocessed to clean and prepare them for classification and, second, they are classified as aggressive or non-aggressive via the classifier. At the end, the number of aggressive comments provided by each expert in every debate round is computed. The value of ζ depends on problem under discussion, i.e., if the effects of the decision adopted are transcendent, the number of debate rounds should be logically high. On the contrary, if the consequences of the decision are still important, but not so transcendental, or it is urgent to adopt the decision, a low number of debate rounds is required.

In the following, we elaborate on in detail both the preprocessing and the classifier constructed.

1) Preprocessing: The preprocessing of the comments provided by the experts is required to transform them into a useful and efficient format for the classifier. It is also necessary for transforming the raw data of both the training data set and the test data set, which are used to build the classifier. Here, as training data set and test data set, we use the data set offered by [47].

There exist different techniques for preprocessing [48]. Among them, we have selected the following ones that allow us to divide the comments, detect what type of words we have, and optimally reduce the bag of words generated: (i) tokenization, which segments the comments into words, or sets of words (it also separates out words and removes punctuation), (ii) stemming and lemmatization, which reduce

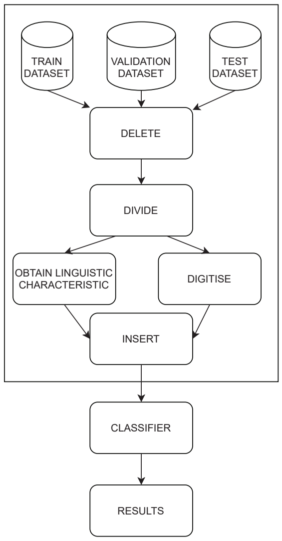

Fig. 2. Outline of the proposed sentiment analysis algorithm.

words to a root by removing inflexion via dropping unneeded characters (usually suffixes), (iii) removal of stop words, which removes commonly used words unlike to be useful for learning, and (iv) part-of-speech tagging (POS-tagging), which consists in assigning parts of speech to every word, such as an adjective, verb, noun, and so on. Using these techniques, the preprocessing is composed of five steps (see Fig. 2):

1) Delete. It consists in simplifying words to their root word and removing the elements that do not give useful information, that is, removing non-relevant words, numbers and double blanks. This is done by using stemming, lemmatization and removal of stop words.
2) Divide. It consists in breaking down the comments into n-grams, that is, the comments are divided into words or sets of words. This is done by using tokenization.
3) Obtain linguistic characteristics. It consists in associating more information with the comments by applying linguistic features. Concretely, the following ones have been used because, after an intensive experimentation, they have shown a good performance to classify comments:
   - Length of the comments. This characteristic represents the number of words composing a comment. It is extracted because an aggressive comment is usually either very long or short and direct.
   - Number of characters. This characteristic determines the total number of characters in the comment. It complements the above characteristic because a comment with a few words but a great number of characters may indicate that the comment is aggressive.
   - Number of negative, neutral and positive emojis. They represent the number of emojis whose predominant value is negative, neutral and positive, respectively. They are extracted because a significant number of negative emojis can determine that the comment is aggressive while a significant number of neutral or positive emojis can determine that the comment is non-aggressive [49].
   - Number of emojis. This characteristic represents the total number of emojis composing a comment. There exists a direct relationship between the total number of emojis and the type of comment because the absence of emojis can determine the type of comment.
   - Number of negative, neutral and positive words. These characteristics determine the number of words whose values in a database are lower, equal and greater than 0, respectively. They are extracted because a large number of negative words can represent an aggressive comment while a great number of neutral or positive words can represent a non-aggressive comment. Refer to [50] to determine the value of a word.
   - Number of words with repeated characters. This characteristic indicates the number of words having consecutive repeated characters. It is used because the increase of repeated characters in a word can determine that the expert is relaxed while writing the comment and, therefore, it can indicate that he or she does not write a comment with aggressive content.
   - Number of consecutive words repeated. It represents the number of consecutive repeated words. Similar to the previous one, a more formal tone is usually used when trying to offend someone and, therefore, fewer word repetitions or misspellings are generated.
   - Number of repeated exclamations. It represents the number of consecutive repeated exclamations. It is chosen because the use of repeated exclamations increases the intensity of the comment and, therefore, it can offer information about the expert's intention.
   - Number of negative and positive classic emoticons. These characteristics represent the number of classic emoticons whose value in a database is negative and positive, respectively. These characteristics complement the emojis as the use of these types of emoticons can determine that a comment is non-aggressive and aggressive, respectively [51].
   - Number of complete words in capital letters. This characteristic represents the total number of words written completely in capital letters. It is extracted because it can determine the intensity to which a comment or a word is expressed. For example, an uppercase comment indicates a very high intensity, which determines that the comment is aggressive.
   - Number of insults. It determines the number of insults composing the comment. A great number of insults indicates that the comment is aggressive.
   - POS-tagging. It allows us to determine the type of each word composing the comment. Using it, we can obtain relevant information about the comment.
4) Digitise. It consists in taking the set of n-grams and digitising them to obtain a numeric matrix. This can be done in three different ways:
   - Binary frequency of terms. It generates a binary numeric matrix as follows: if the word or set of words is found in the comment, a value of 1 is placed; otherwise, a value of 0 is set.

- Frequency of terms. It works similarly to the previous one. However, instead of placing 1 if the word or set of words is found in the comment, it places the number of times that the word or the set of words are found in the comment. Therefore, it generates a numeric matrix containing the frequency of terms.
- Term frequency-inverse document frequency (TF-IDF). It intends to reflect how important a word is to a document in a corpus or collection. Therefore, here, it aims to weigh the importance of a word to comment on in the debate. It generates a numeric matrix containing the TF-IDF.

5) Insert. Once the linguistic characteristics have been obtained, they are added to the numeric matrix. Finally, another column is added to the matrix, which represents whether the comment is aggressive (a value of 1 is then added) or non-aggressive (a value of 0 is placed).

2) Classifier: Several machine learning algorithms can be used to construct the classifier based on supervised learning for sentiment analysis [46]. Among them, we have chosen the following ones, which are not based on deep learning and, as a result, allow us a faster training:

- SVM [52]. It tries to discover the best hyperplane in more than two dimensions, or the best line in two dimensions, to separate the space into classes. The line (hyperplane) is obtained by means of the maximum margin, i.e., the maximum distance between data points of both classes.
- Naïve Bayes [53]. The crux of this algorithm is based on the Bayes theorem and in the assumption that the predictor variables are independent, i.e., the presence of a particular predictor variable does not affect the others.
- KNN [54]. It is based on the principle that instances within a data set usually exist close to others having similar properties. If the instances are tagged with a label, then the label of an unclassified instance can be obtained by observing the class of its k nearest neighbours and assigning to it the most frequent class label.

A classifier can be built based on each algorithm. First, we must adjust its parameters via a training data set (as mentioned, the data set offered by [47] is used as a training data set). To do so, the training data set must be preprocessed as described in Section III-A1 to obtain a numeric matrix that is used as input for the algorithm [55]. Second, once the classifier has been built, we must evaluate its quality by means of a test data set (the data set offered by [47] is also used as test data set), which must be preprocessed too.

The best accuracy, i.e., the percentage of words classified correctly as aggressive and non-aggressive, reached by each algorithm is shown in Fig. 3. This best accuracy is achieved by using these combinations:

- SVM. The best combination is based on the use of stemming and removal of stop words, all the linguistic characteristics, unigrams, and the TF-IDF matrix. Using this combination, the SVM reaches an accuracy of 83%.
- Naïve Bayes. The best combination is based on the use of stemming and removal of stop words, all the linguistic characteristics, unigrams, and the binary matrix. Using this combination, the Naïve Bayes algorithm reaches an accuracy of 80%.
- KNN. The best combination is based on the use of stemming and removal of stop words, unigrams, the frequency matrix, a value of 4 for k, and without linguistic characteristics. Using this combination, the KNN algorithm reaches an accuracy of 74%.

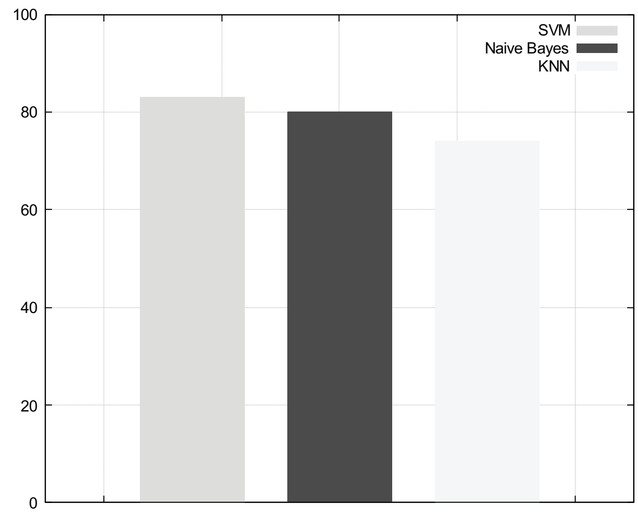

Fig. 3. Best accuracy achieved by SVM, Naïve Bayes, and KNN.

According to these results, the classifier based on the SVM algorithm will be used to identify aggressive comments during the debate as it has achieved the best accuracy. In addition, in order to not be only based on the data set offered by [47], this classifier is reinforced with other dictionaries, such as the insult dictionary, and other data sets measuring emotions. In this way, the classifier does not rely solely on the comments contained in the training data set.

B. Providing Assessments

Once the debate is closed, the experts must provide their assessments. As mentioned in Section II-B, we can find both different representation formats and distinct representation domains to model these assessments [35]. Among them, in this study, we assume preference relations as representation format and [0,1]-values as representation domain, i.e., we assume fuzzy preference relations due to they have been widely and successfully used in GDM [56].

The fuzzy preference relation P^z provided by the expert e_z on a collection of alternatives X is modelled by a function $\mu_{P^z} : X \times X \rightarrow [0,1]$. To represent the fuzzy preference relation P^z given by the expert e_z, a matrix $P^z = (p^z_{jk})$ of size $n \times n$ is generally utilized, being $p^z_{jk} = \mu_{P^z}(x_j, x_k)$ the preference degree of x_j over x_k $(j \neq k)$ according to e_z. Concretely, whether $p^z_{jk} > 0.5$, then x_j is preferred over x_k by e_z; whether $p^z_{jk} < 0.5$, then x_k is preferred over x_j by e_z; and whether $p^z_{jk} = 0.5$, then both x_j and x_k are equally preferred or deemed as indifferent by e_z. Therefore, assuming a group of experts, $E = \{e_1, e_2, \ldots, e_m\}$, and a collection of alternatives, $X = \{x_1, x_2, \ldots, x_n\}$, in this stage, we obtain a set of m fuzzy preference relations, P^z, $z = 1, \ldots, m$.

C. Calculating Experts' Weights

The experts' behaviour during the debate is considered to assign weights to them. That is, we know that an aggressive behaviour leads to the deterioration of the debate and hinders the achievement of consensus [13], [14]. Therefore, lower weights should be assigned to experts using an aggressive language during the debate.

Two different methods are developed, which use the results returned by the classifier: (i) a global method, and (ii) an iterative method.

- Global method. First, we calculate the total number of aggressive comments, $ac_z \in \mathbb{N} \cup \{0\}; ac_z \leq tc_z; z = 1, \ldots, m$, and the total number of comments provided, $tc_z \in \mathbb{N}; z = 1, \ldots, m$, i.e., both aggressive and non-aggressive comments, expressed by the expert e_z during all the debate rounds. They are computed as follows:

$$ac_z = \sum_{j=1}^{\zeta} ac_z^j \tag{1}$$

$$tc_z = \sum_{j=1}^{\zeta} tc_z^j \tag{2}$$

  being $ac_z^j \in \mathbb{N} \cup \{0\}; ac_z^j \leq tc_z^j$ the number of aggressive comments given by the expert e_z in the round j and $tc_z^j \in \mathbb{N}$ the number of comments, both aggressive and non-aggressive comments, provided by the expert e_z in the round j.
  Second, a peaceful coefficient, pc_z, assuming values in-between $[0, 1]$ is computed for each expert e_z as follows:

$$pc_z = 1 - \frac{ac_z}{tc_z}. \tag{3}$$

  The lower the value of pc_z, the greater the aggressiveness of e_z.
  Third, the weight, w_z^g, associated with each expert e_z, being $\sum_{z=1}^{m} w_z^g = 1$, is determined by dividing her or his peaceful coefficient, pc_z, by the total sum of all the peaceful coefficients of the experts:

$$w_z^g = \frac{pc_z}{\sum_{z=1}^{m} pc_z}. \tag{4}$$

- Iterative method. Unlike the global method, it uses the weights obtained in previous debate rounds to compute the weight of the current round. This method starts initializing the weight associated with each expert, w_z^0, with a value equal to $\frac{1}{m}$, being m the number of experts. Then, a peaceful coefficient, pc_z^j, assuming values in-between $[0, 1]$ is computed for each expert e_z. Unlike the peaceful coefficient calculated by the global method, it is calculated according to the number of aggressive comments and the number of comments, both aggressive and non-aggressive comments, provided in the round j:

$$pc_z^j = 1 - \frac{ac_z^j}{tc_z^j}. \tag{5}$$

  Using the peaceful coefficient computed in each round, the weight associated with each expert, e_z, in the debate around j is calculated as follows:

$$w_z^j = \frac{pc_z^j \cdot w_z^{j-1}}{\sum_{z=1}^{m} pc_z^j \cdot w_z^{j-1}}. \tag{6}$$

  In the case that all experts have a weight equal to 0 since all of them have provided only aggressive comments in a given round, the weights are reset to $\frac{1}{m}$. Finally, the weight, w_z^i, associated with each expert e_z is determined by the weight reached in the last round, i.e.:

$$w_z^i = w_z^{\zeta}. \tag{7}$$

D. Analyzing Consensus

Motivated by the possibility that some experts might exhibit disagreements with each other's assessments or they may sometimes not accept the decision made, consensus reaching processes were introduced as a very important requirement in GDM processes to reach a high level of collective agreement before making the decision [10]. Reaching consensus needs experts to change their first assessments, bringing them closer to each other, towards a collective assessment deemed as good enough by the group. Its principal steps are: (i) consensus measurement, i.e., determining the current closeness level between assessments via a consensus measure [57], (ii) consensus control, i.e., determining whether the required consensus level has been reached, and (iii) reconsideration of judgments, i.e., debating again and modifying assessments to bring the furthest assessments closer to the remainder of the group members' assessments to increase the consensus.

To determine the consensus achieved, we introduce two novel approaches considering the experts' behaviour during the debate:

- An approach based on the standardised Euclidean distance, called SED, which makes use of the Euclidean distance between the experts' assessments and the experts' weights to compute the agreement between two experts. Because there exist two methods to compute the weights, by applying this approach we obtain two consensus measures, $CGSED_y^z$ and $CISED_y^z$, measuring the agreement between the experts, e_z and e_y, when using the global method and the iterative method, respectively, to compute the experts' weights:

$$CGSED_y^z = 1 - \frac{\sum_{j=1}^{n} \sum_{k=1; k \neq j}^{n} \sqrt{(p_{jk}^z - p_{jk}^y)^2 + (w_z^g - w_y^g)^2}}{(n^2 - n) \cdot \sqrt{2}} \tag{8}$$

$$CISED_y^z = 1 - \frac{\sum_{j=1}^{n} \sum_{k=1; j \neq j}^{n} \sqrt{(p_{jk}^z - p_{jk}^y)^2 + (w_z^i - w_y^i)^2}}{(n^2 - n) \cdot \sqrt{2}}. \tag{9}$$

Once the consensus measures between every pair of experts have been computed, two global consensus measures, $CGSED$ and $CISED$ are computed:

$$CGSED = 2 \cdot \frac{\sum_{z=1}^{m-1} \sum_{y=z+1}^{m} CGSED_y^z}{m^2 - m} \tag{10}$$

$$CISED = 2 \cdot \frac{\sum_{z=1}^{m-1} \sum_{y=z+1}^{m} CISED_y^z}{m^2 - m}. \tag{11}$$

- An approach based on the product of the consensus between the experts' assessments and the consensus between the experts' weights. It assumes that experts should have both similar assessments and a similar degree of aggressiveness, which is determined by the weights. As before, by applying this approach, we obtain two consensus measures, CGP_y^z and CGP_y^z, measuring the agreement between the experts, e_z and e_y, when using the global method and the iterative method, respectively, to compute the experts' weights:

$$CGP_y^z = \left(1 - \frac{\sum_{j=1}^{n} \sum_{k=1;k\neq j}^{n} |p_{jk}^z - p_{jk}^y|}{(n^2 - n)} \right) \cdot (1 - |w_z^g - w_y^g|) \tag{12}$$

$$CIP_y^z = \left(1 - \frac{\sum_{j=1}^{n} \sum_{k=1;k\neq j}^{n} |p_{jk}^z - p_{jk}^y|}{(n^2 - n)} \right) \cdot (1 - |w_z^i - w_y^i|). \tag{13}$$

Once the consensus measures between every pair of experts have been computed, two global consensus measures, CGP and CIP are computed as follows:

$$CGP = 2 \cdot \frac{\sum_{z=1}^{m-1} \sum_{y=z+1}^{m} CGP_y^z}{m^2 - m} \tag{14}$$

$$CIP = 2 \cdot \frac{\sum_{z=1}^{m-1} \sum_{y=z+1}^{m} CIP_y^z}{m^2 - m}. \tag{15}$$

Once the consensus among the experts has been measured, we must check whether it is enough. To do so, a consensus threshold $\alpha \in [0, 1]$ is established, which depends on the problem [10]. Whether the consensus achieved is equal to or greater than the consensus threshold, we can obtain the ranking of the alternatives derived from the collective assessment. Otherwise, the experts must reconsider their assessments by starting a new discussion round, i.e., the first stage is carried out again. A limited number of discussion rounds, $\beta \in \mathbb{N}$, is established to avoid a never-ending decision-making process. When this limit is achieved, we proceed to obtain the ranking of the alternatives independently of the agreement achieved.

E. Obtaining the Collective Assessment

To compute the collective assessment, all the individual assessments must be aggregated. Here, as we are modelling the assessments via fuzzy preference relations, we must fuse all of them to obtain a collective fuzzy preference relation by means of an aggregation function [41].

To consider the experts' behaviour during the debate stage in the calculation of the collective assessment, the weights associated with the experts are used to obtain the fuzzy collective preference relation. Concretely, the weighted mean is used as an aggregation function. However, other aggregation functions such as the power average operator or the Bonferroni mean-type aggregation functions could be considered [58], [59]. As two methods can be applied to compute the weights, two collective fuzzy preference relations can be obtained:

- A collective fuzzy preference relation, $C^g = (c_{jk}^g)$, using the global method is obtained as follows:

$$c_{jk}^g = \sum_{z=1}^{m} w_z^g \cdot p_{jk}^z. \tag{16}$$

- A collective fuzzy preference relation, $C^i = (c_{jk}^i)$, using the iterative method is obtained as follows:

$$c_{jk}^i = \sum_{z=1}^{m} w_z^i \cdot p_{jk}^z. \tag{17}$$

Once the collective fuzzy preference relations have been computed, the ranking of alternatives can be derived from them. This is done in the next stage.

F. Ranking the Alternatives

In this stage, the information contained in the collective fuzzy preference relation is exploited to rank the alternatives from best to worst as solutions to the decision-making process. Among the different functions that can be applied to do it [42], [60], we use both the quantifier-guided dominance degree and the quantifier-guided non-dominance degree:

- The quantifier-guided dominance degrees, $QGDD_j^g$ and $QGDD_j^i$, determine the dominance that x_j has over the other alternatives when using the collective fuzzy preference relations, C^g and C^i, respectively. They are computed as:

$$QGDD_j^g = \frac{1}{n-1} \sum_{k=1;k\neq j}^{n} c_{jk}^g \tag{18}$$

$$QGDD_j^i = \frac{1}{n-1} \sum_{k=1;k\neq j}^{n} c_{jk}^i. \tag{19}$$

- The quantifier-guided non-dominance degrees, $QGNDD_j^g$ and $QGNDD_j^i$, determine the degree in which x_j is not dominated by the other alternatives when using the collective fuzzy preference relations, C^g and C^i, respectively. They are computed as:

$$QGNDD_j^g = \frac{1}{n-1} \sum_{k=1;k\neq j}^{n} 1 - d_{kj}^g \tag{20}$$

$$QGNDD_j^i = \frac{1}{n-1} \sum_{k=1; k \neq j}^{n} 1 - d_{kj}^i \quad (21)$$

being $d_{kj}^g = \max\{c_{kj}^g - c_{jk}^g, 0\}$ and $d_{kj}^i = \max\{c_{kj}^i - c_{jk}^i, 0\}$, respectively, which determine the degree in which the alternative x_j is dominated by the alternative x_k.

The application of these choice degrees over the set of alternatives X is performed according to the following approach:

- The application of $QGDD_j^g$ and $QGDD_j^i$ over X to get the following sets of alternatives:

$$X^{QGDD^g} = \{x_j \in X \mid QGDD_j^g = \sup_{x_k \in X} QGDD_k^g\} \quad (22)$$

$$X^{QGDD^i} = \{x_j \in X \mid QGDD_j^i = \sup_{x_k \in X} QGDD_k^i\}. \quad (23)$$

- The application of the intersection to the above sets to get the following set of alternatives:

$$X^{QGDD} = X^{QGDD^g} \bigcap X^{QGDD^i}. \quad (24)$$

If $\#X^{QGDD} = 1$, then this is an alternative chosen as a solution to the problem. Otherwise, continue.
- The application of $QGNDD_j^g$ and $QGNDD_j^i$ over X^{QGDD} to get the following set of alternatives:

$$X^{QGNDD} = \{x_j \in X \mid QGNDD_j = \sup_{x_k \in X} QGNDD_k\} \quad (25)$$

where

$$QGNDD_k = \frac{QGNDD_k^g + QGNDD_k^i}{2}. \quad (26)$$

This is the selection set of alternatives.

IV. Case Study: Profitable Product Selection

This section conducts a real-world problem in which an enterprise is looking at changing its solar panel product line for the next year. It wants to carry only one brand of solar panels. It has received four offers, $X = \{x_1, x_2, x_3, x_4\}$, from different solar panel manufacturers. The enterprise consults four experts, $E = \{e_1, e_2, e_3, e_4\}$, about the brand that could offer the highest total profit. Specific enterprise, brands, and experts, are omitted for privacy reasons.

First, the experts' behaviour during the debate is determined. For this proposal, in every debate round, the experts discuss the alternatives and the classifier described in Section III-A is applied to detect the aggressive language. Here, the number of debate rounds, ζ, is equal to 5 (4 rounds to compare options and 1 round to draw conclusions). In Table I, the results returned by the classifier are shown. Due to space limitations, we do not list the comments provided during the debate rounds in this paper, but we upload it as supplementary material and available at http://ieee-dataport.org/10589.

Table I
Number of Aggressive Comments (ac) and Total Comments (tc) Provided by the Experts in Every Debate Round

Round	ac_1	tc_1	ac_2	tc_2	ac_3	tc_3	ac_4	tc_4
1	2	10	1	14	3	11	1	15
2	3	11	2	11	5	12	2	14
3	0	10	1	13	3	13	1	10
4	3	15	1	10	4	10	0	10
5	1	13	4	13	1	11	1	11

Second, and once the debate stage has finished and the aggressive comments have been detected, the experts provide their assessments:

$$P^1 = \begin{pmatrix} - & 0.8 & 0.8 & 0.9 \\ 0.2 & - & 0.5 & 0.6 \\ 0.1 & 0.5 & - & 0.6 \\ 0.2 & 0.3 & 0.4 & - \end{pmatrix} P^2 = \begin{pmatrix} - & 0.2 & 0.5 & 0.4 \\ 0.8 & - & 0.9 & 0.8 \\ 0.4 & 0.2 & - & 0.5 \\ 0.5 & 0.1 & 0.5 & - \end{pmatrix}$$

$$P^3 = \begin{pmatrix} - & 0.5 & 0.0 & 0.3 \\ 0.6 & - & 0.1 & 0.4 \\ 1.0 & 0.8 & - & 0.9 \\ 0.6 & 0.5 & 0.2 & - \end{pmatrix} P^4 = \begin{pmatrix} - & 0.5 & 0.3 & 0.1 \\ 0.5 & - & 0.5 & 0.2 \\ 0.6 & 0.4 & - & 0.0 \\ 0.8 & 0.9 & 1.0 & - \end{pmatrix}$$

Third, based on the aggressive language used during the debate, the weights associated with the experts are obtained to take them into account in the rest of the decision-making process. Using (4) and (7), the weights are calculated according to the global method and the iterative method, respectively:

$$\begin{array}{llll} w_1^g = 0.2540 & w_2^g = 0.2555 & w_3^g = 0.2156 & w_4^g = 0.2748 \\ w_1^i = 0.2529 & w_2^i = 0.2572 & w_3^i = 0.1048 & w_4^i = 0.3852 \end{array}$$

As it can be seen, the iterative and the global methods assign similar weights to the experts e_1 and e_2. However, the weights assigned to the experts e_3 and e_4 are very different. This is because the iterative method uses the weight obtained in the previous debate round whereas the global method makes a global calculation. In such a way, it could be the case that all comments provided by an expert in a debate round are aggressive. Then, the global method would decrease the weight assigned to the expert whereas the iterative method would cancel his or her weight out during the whole debate.

Fourth, the consensus achieved is measured and compared with the consensus threshold, α, established, which is 0.7 in this problem. Using (8), (9), (12) and (13), the consensus measures $CGSED_y^z$, $CISED_y^z$, CGP_y^z and CIP_y^z, measuring the agreement between the experts e_z and e_y, are calculated:

$$\begin{array}{llllll} CGSED_2^1 & = & 0.77 & CISED_2^1 & = & 0.77 \\ CGP_2^1 & = & 0.67 & CIP_2^1 & = & 0.67 \\ \\ CGSED_3^1 & = & 0.71 & CISED_3^1 & = & 0.69 \\ CGP_3^1 & = & 0.57 & CIP_3^1 & = & 0.50 \\ \\ CGSED_4^1 & = & 0.69 & CISED_4^1 & = & 0.67 \\ CGP_4^1 & = & 0.55 & CIP_4^1 & = & 0.48 \end{array}$$

$$\begin{array}{lcllcl} CGSED_3^2 & = & 0.72 & CISED_3^2 & = & 0.69 \\ CGP_3^2 & = & 0.58 & CIP_3^2 & = & 0.51 \end{array}$$

$$\begin{array}{lcllcl} CGSED_4^2 & = & 0.73 & CISED_4^2 & = & 0.71 \\ CGP_4^2 & = & 0.60 & CIP_4^2 & = & 0.54 \end{array}$$

$$\begin{array}{lcllcl} CGSED_4^3 & = & 0.74 & CISED_4^3 & = & 0.66 \\ CGP_4^3 & = & 0.60 & CIP_4^3 & = & 0.46 \end{array}$$

Then, the global consensus measures, $CGSED$, $CISED$, GCP and CIP, are computed by means of (10), (11), (14) and (15), respectively:

$$\begin{array}{lcllcl} CGSED & = & 0.72 & CISED & = & 0.70 \\ CGP & = & 0.59 & CIP & = & 0.53 \end{array}$$

Due to two of the global consensus measures are lower than α, the experts must discuss again and reconsider their assessments to increase the consensus. Therefore, a new discussion round (the second one) is carried out (we suppose the maximum number of discussion round, β, is equal to 5 in this problem). In Table II, the results returned by the classifier in the second debate round are shown (check http://ieee-dataport.org/10589 to see the comments provided by the experts during this debate round).

Once the debate has finished and the aggressive comments have been detected, the experts provide their new assessments, which are:

$$P^1 = \begin{pmatrix} - & 0.5 & 0.4 & 0.9 \\ 0.8 & - & 0.0 & 0.6 \\ 1.0 & 0.9 & - & 1.0 \\ 0.0 & 0.2 & 0.1 & - \end{pmatrix} P^2 = \begin{pmatrix} - & 0.2 & 0.1 & 0.6 \\ 1.0 & - & 0.3 & 0.7 \\ 0.9 & 0.9 & - & 0.8 \\ 0.2 & 0.2 & 0.5 & - \end{pmatrix}$$

$$P^3 = \begin{pmatrix} - & 0.5 & 0.0 & 0.3 \\ 0.6 & - & 0.1 & 0.4 \\ 1.0 & 0.8 & - & 0.9 \\ 0.6 & 0.5 & 0.2 & - \end{pmatrix} P^4 = \begin{pmatrix} - & 0.4 & 0.4 & 0.9 \\ 0.9 & - & 0.4 & 0.8 \\ 0.8 & 0.8 & - & 0.9 \\ 0.3 & 0.2 & 0.2 & - \end{pmatrix}$$

It can be observed that expert e_3 has decided to keep their initial assessments and, therefore, e_3 provides the same fuzzy preference relation.

Using the results provided by the classifier in this second discussion round, the weights associated with the expert are recalculated:

$$\begin{array}{llll} w_1^g = 0.2613 & w_2^g = 0.2302 & w_3^g = 0.2575 & w_4^g = 0.2510 \\ w_1^i = 0.3058 & w_2^i = 0.1468 & w_3^i = 0.2892 & w_4^i = 0.2583 \end{array}$$

Table II
NUMBER OF AGGRESSIVE COMMENTS (ac) AND TOTAL COMMENTS (tc) PROVIDED BY THE EXPERTS IN EVERY DEBATE ROUND

Round	ac_1	tc_1	ac_2	tc_2	ac_3	tc_3	ac_4	tc_4
1	0	14	2	12	2	14	3	15
2	3	12	3	11	4	14	0	12
3	0	14	1	13	1	12	0	12
4	2	14	5	11	1	14	3	15
5	2	14	2	15	0	15	3	11

Using these new weights, the consensus measures $CGSED_y^z$, $CISED_y^z$, CGP_y^z and CIP_y^z, are computed again:

$$\begin{array}{lcllcl} CGSED_2^1 & = & 0.85 & CISED_2^1 & = & 0.81 \\ CGP_2^1 & = & 0.78 & CIP_2^1 & = & 0.67 \end{array}$$

$$\begin{array}{lcllcl} CGSED_3^1 & = & 0.83 & CISED_3^1 & = & 0.83 \\ CGP_3^1 & = & 0.76 & CIP_3^1 & = & 0.75 \end{array}$$

$$\begin{array}{lcllcl} CGSED_4^1 & = & 0.89 & CISED_4^1 & = & 0.90 \\ CGP_4^1 & = & 0.86 & CIP_4^1 & = & 0.83 \end{array}$$

$$\begin{array}{lcllcl} CGSED_3^2 & = & 0.79 & CISED_3^2 & = & 0.82 \\ CGP_3^2 & = & 0.73 & CIP_3^2 & = & 0.64 \end{array}$$

$$\begin{array}{lcllcl} CGSED_4^2 & = & 0.86 & CISED_4^2 & = & 0.89 \\ CGP_4^2 & = & 0.83 & CIP_4^2 & = & 0.76 \end{array}$$

$$\begin{array}{lcllcl} CGSED_4^3 & = & 0.82 & CISED_4^3 & = & 0.82 \\ CGP_4^3 & = & 0.75 & CIP_4^3 & = & 0.73 \end{array}$$

Then, the global consensus measures, $CGSED$, $CISED$, GCP and CIP, are:

$$\begin{array}{lcllcl} CGSED & = & 0.85 & CISED & = & 0.83 \\ CGP & = & 0.78 & CIP & = & 0.73 \end{array}$$

As all the global consensus measures are equal or higher than $\alpha = 0.7$, we can proceed to obtain the ranking of the alternatives. It means the collective fuzzy preference relations, C^g and C^i, are calculated according to (16) and (17), respectively:

$$C^g = \begin{pmatrix} - & 0.4316 & 0.2279 & 0.6764 \\ 0.8196 & - & 0.1952 & 0.6217 \\ 0.9268 & 0.8492 & - & 0.9031 \\ 0.2758 & 0.2773 & 0.2429 & - \end{pmatrix}$$

$$C^i = \begin{pmatrix} - & 0.4591 & 0.2403 & 0.6825 \\ 0.7973 & - & 0.1763 & 0.6085 \\ 0.9337 & 0.8453 & - & 0.9159 \\ 0.2803 & 0.2868 & 0.2134 & - \end{pmatrix}$$

Using both collective fuzzy preference relations, the ranking of alternatives is generated. To do so, the quantifier-guided dominance degrees and the quantifier-guided non-dominance degrees are computed using (18)–(21):

$$\begin{array}{lcllcl} QGDD_1^g & = & 0.4606 & QGNDD_1^g & = & 0.6561 \\ QGDD_1^i & = & 0.4453 & QGNDD_1^i & = & 0.6377 \\ QGDD_2^g & = & 0.5274 & QGNDD_2^g & = & 0.7770 \\ QGDD_2^i & = & 0.5455 & QGNDD_2^i & = & 0.7820 \\ QGDD_3^g & = & 0.8983 & QGNDD_3^g & = & 1.0000 \\ QGDD_3^i & = & 0.8930 & QGNDD_3^i & = & 1.0000 \\ QGDD_4^g & = & 0.2602 & QGNDD_4^g & = & 0.5246 \\ QGDD_4^i & = & 0.2653 & QGNDD_4^i & = & 0.5316 \end{array}$$

Using these choice degrees, the following sets of alternatives are obtained:

$$X^{QGDD^g} = \{x_3\}$$

$$X^{QGDD^i} = \{x_3\}$$

Hence, $X^{QGDD} = \{x_3\}$ and x_3 is chosen as the solution to the decision-making problem according to the experts' assessments, i.e., x_3 must be selected as panel solar brand.

V. Concluding Remarks

In this study, we have presented a new GDM method based on the experts' behaviour during the debate stage carried out before providing the assessments. The existing GDM methods do not take into account the experts' behaviour during the debate stage in the remainder of the decision-making process and, by not considering it, information that could be of interest is lost. However, this new GDM method incorporates a classifier based on sentiment analysis techniques that can extract additional data generated during the debate stage (i.e., the experts' behaviour) and take advantage of it. Concretely, based on the experts' behaviour, two procedures have been developed to assign weights to the experts and two new consensus measures have been proposed.

Whether the models for GDM deal with different aspects of the decision-making process, a comparison of the results returned by a model with others is not a straightforward task. The features considered by the models are different and, as a consequence, a quantitative comparison would not be meaningful. In any case, in the following, we analyse some advantages and shortcomings of the proposed approach:

- By incorporating a classifier, which is based on sentiment analysis techniques and makes use of different linguistic features, the proposed model can categorize the experts' behaviour during the debate stage. In comparison with the existing methods based on sentiment analysis techniques that analyse the debate conducted between the experts [15]–[19], it is a notable advantage because it has been demonstrated that, first, an aggressive language negatively affects the trustworthiness of the speaker and her or his credibility [13] and, second, it also leads to the deterioration of the debate [14]. Both scenarios must be avoided because they make difficult the achievement of consensus, which is the most important point in GDM processes [10]. Therefore, the model proposed in this study, by exploiting the information returned by the sentiment analysis classifier, can better model the consensus reaching process carried out in a GDM process.
- Related to the above advantage, it introduces two new procedures assigning weights to the experts according to their behaviours during the debate stage. This allows considering the experts' behaviour in the remainder of the GDM process, specifically, in the analysis of the consensus (when calculating the consensus achieved between the experts' assessments) and in the computation of the collective assessment (when the individual experts' assessments are aggregated). On the one hand, the procedure calculating the overall weight obtains an average weighting of the aggressive comments provided by an expert. Its advantage is that even if all the comments provided in a round are aggressive, the weight of the expert can be compensated in other rounds where any aggressive comment is provided. Nevertheless, it has the disadvantage that it is more general, which implies that the expert's behaviour can only be seen during the debate as a whole and not in each round. On the other hand, the procedure calculating the iterative weight uses the weight of the previous round to calculate the new one. It has the advantage that it can be analyzed in each round in a more detailed way because the weight related to the expert is altered according to the previous round. However, it is stricter because in the case that all the comments provided by an expert are aggressive, the weight of this expert will be equal to zero, although, in the other rounds, the expert does not provide any aggressive comments. Then, by using these new two procedures, two novel consensus measures and two novel methods to compute the collective assessment that consider the experts' behaviour during the debate stage are proposed.
- The classifier based on sentiment analysis techniques uses different linguistic characteristics that provide relevant information when classifying the comments. It presents the following advantages. First, it detects aggressive comments by using different linguistic characteristics that complete the preprocessing. Some of them have already been used, such as the use of POS-tagging, but others are new, such as the number of complete words written in capital letters or the consecutive number of repeated exclamations, to cite some examples. Second, it uses an optimised preprocessing matrix built via different methods to remove words that do not provide useful information and to reduce words to their word stems (it reduces the number of elements in the bag of words). However, its shortcoming is that it could not find out all the linguistic characteristics that are present. As a result, the numeric matrix would be less complete and this could generate some false negatives, i.e., an aggressive comment could be classified as a non-aggressive comment.
- Consensus measures. Usually, the consensus is determined by looking at the difference between the experts' assessments. However, this is incomplete as the experts' behaviour can be as relevant as the assessments. For instance, two experts can agree that a given alternative is better than other one, but the reasons given by each expert can be different. Consequently, both experts have the same opinion, but the arguments considered to give that opinion are different. For this reason, the consensus measures proposed in this study do only consider the assessments but also the behaviour shown to provide them. It makes the consensus measures more complete than if only the experts' assessments are considered.

The number of GDM models using sentiment analysis has grown in the last years. In [18], the authors present a method based on sentiment analysis and apply it to decision-making. The method creates two bags of words from its dictionaries. Nevertheless, it creates a bag of words from pre-classified comments and uses a ranking algorithm to classify the comments made by experts. Moreover, it applies that ranking, influencing the weights of the experts. In [19], the authors introduce a method applicable to a large number of alternatives that uses sentiment analysis to find out which alternative is better in each round. However, our method uses two procedures that allow us to check who expert is more aggressive, in general, and in each round, during the debate and that could be applied to each alternative in future works. Thus, obtaining a classifier that analyses the likes/dislikes of each expert on an alternative. Furthermore, the method presented in this paper uses a previously defined bag of words and does not use a classifier while ours uses a classifier, such as a SVM and uses a known data set containing comments. Finally, in [16], the authors develop a method that can analyse the debate in social networks by using sentiment analysis. Nonetheless, the method presented in this paper does not only analyses the debate by using sentiment analysis but also applies it to the GDM model by adjusting the weights in two different ways. This aims to have a double analysis of the debate, one in a general way and the other in an iterative way.

This research may be continued as follows. First, as we have mentioned, the sentiment analysis classifier could ignore some linguistic characteristics, classifying a comment in a wrong way. Therefore, new linguistic characteristics should be investigated and verified if they are useful or not to classify comments as aggressive or non-aggressive. Second, social networks have changed the scenario in which GDM processes are carried out. They have facilitated that people participate in the decision-making processes, which has increased the number of individuals trying to solve a GDM problem. It has given rise to a new research area within GDM, called large-scale GDM [61], that has gained great attention in recent years. For instance, sustainable building material selection and bid evaluation are two examples of real-life scenarios in which large-scale GDM methods have been applied [62], [63]. The existing methods dealing with large-scale GDM usually divide the experts into groups according to their assessments, i.e., experts providing similar assessments are in the same group. However, other criteria could be also used to group the experts. For example, the behaviour during the debate stage could be considered to classify experts with similar behaviour in the same group. In such a case, an optimization process could be required in the process of aggressiveness detection due to the high amount of information generated. And third, to help experts improve the consensus in successive debate rounds, a feedback mechanism could be incorporated to advise experts on how they should modify their assessments if they want to make their positions converge, and, consequently, to improve the consensus [64].

References

[1] A. Dorison, J. Klusowski, S. Han, and J. S. Lerner, "Emotion in organizational judgment and decision making," *Organ. Dyn.*, vol. 49, no. 1, Jan.–Mar. 2020, Art. no. 100702.

[2] H. A. Simon, "Making management decisions: The role of intuition and emotion," *Acad. Manage. Exec.*, vol. 1, no. 1, pp. 57–64, Feb. 1987.

[3] A. V. Banerjee, "A simple model of herd behavior," *Q. J. Econ.*, vol. 107, no. 3, pp. 797–817, Aug. 1992.

[4] M. Zeelenberg, W. W. van Dijk, A. S. R. Manstead, and J. vanr de Pligt, "On bad decisions and disconfirmed expectancies: The psychology of regret and disappointment," *Cogn. Emot.*, vol. 14, no. 4, pp. 521–541, 2000.

[5] C.-L. Hwang and M.-J. Lin, *Group Decision Making Under Multiple Criteria: Methods and Applications*. Berlin Heidelberg: Springer-Verlag, 1987.

[6] J. A. Morente-Molinera, G. Kou, K. Samuylov, F. J. Cabrerizo, and E. Herrera-Viedma, "Using argumentation in expert's debate to analyze multi-criteria group decision making method results," *Inf. Sci.*, vol. 573, pp. 433–452, Sep. 2021.

[7] S. Saint and J. R. Lawson, *Rules for Reaching Consensus: A Modern Approach to Decision Making*. Jossey-Bass, 1994.

[8] C. T. Butler and A. Rothstein, *On Conflict and Consensus: A Handbook on Formal Consensus Decision Making*. Takoma Park, WA: Food Not Bombs Publishing, 2006.

[9] F. J. Cabrerizo, F. Chiclana, R. Al-Hmouz, A. Morfeq, A. S. Balamash, and E. Herrera-Viedma, "Fuzzy decision making and consensus: Challenges," *J. Intell. Fuzzy Syst.*, vol. 29, no. 3, pp. 1109–1118, 2015.

[10] E. Herrera-Viedma, F. J. Cabrerizo, J. Kacprzyk, and W. Pedrycz, "A review of soft consensus models in a fuzzy environment," *Inf. Fusion*, vol. 17, pp. 4–13, May 2014.

[11] H. Zhang, S. Zhao, G. Kou, C.-C. Li, Y. C. Dong, and F. Herrera, "An overview on feedback mechanisms with minimum adjustment or cost in consensus reaching in group decision making: Research paradigms and challenges," *Inf. Fusion*, vol. 60, pp. 65–79, Aug. 2020.

[12] L. Amgoud and H. Prade, "Using arguments for making and explaining decisions," *Artif. Intell.*, vol. 173, no. 3–4, pp. 413–436, Mar. 2009.

[13] L. König and R. Jucks, "Hot topics in science communication: Aggressive language decreases trustworthiness and credibility in scientific debates," *Public Underst. Sci.*, vol. 28, no. 4, pp. 401–416, 2019.

[14] K. D. Vohs, R. F. Baumeister, and G. Loewenstein, *Do Emotions Help or Hurt Decision Making?: A Hedgefoxian Perspective*. Russell Sage Foundation, 2007.

[15] Z.-S. Chen, X.-L. Liu, K.-S. Chin, W. Pedrycz, K.-L. Tsui, and M. J. Skibniewski, "Online-review analysis based large-scale group decision-making for determining passenger demands and evaluating passenger satisfaction: Case study of high-speed rail system in China," *Inf. Fusion*, vol. 69, pp. 22–39, May 2021.

[16] J. A. Morente-Molinera, G. Kou, Y. Peng, C. Torres-Albero, and E. Herrera-Viedma, "Analysing discussions in social networks using group decision making methods and sentiment analysis," *Inf. Sci.*, vol. 447, pp. 157–168, Jun. 2018.

[17] J. A. Morente-Molinera, G. Kou, C. Pang, F. J. Cabrerizo, and E. Herrera-Viedma, "An automatic procedure to create fuzzy ontologies from users' opinions using sentiment analysis procedures and multi-granular fuzzy linguistic modelling methods," *Inf. Sci.*, vol. 476, pp. 222–238, Feb. 2019.

[18] J. A. Morente-Molinera, G. Kou, K. Samuylov, R. Ureña, and E. Herrera-Viedma, "Carrying out consensual group decision making processes under social networks using sentiment analysis over comparative expressions," *Knowledge-Based Syst.*, vol. 165, pp. 335–345, Feb. 2019.

[19] J. A. Morente-Molinera, F. J. Cabrerizo, J. Mezei, C. Carlsson, and E. Herrera-Viedma, "A dynamic group decision making process for high number of alternatives using hesitant fuzzy ontologies and sentiment analysis," *Knowledge-Based Syst.*, vol. 195, May 2020, Art. no. 105657.

[20] X. Gou, Z. S. Xu, H. C. Liao, and F. Herrera, "Consensus model handling minority opinions and noncooperative behaviors in large-scale group decision-making under double hierarchy linguistic preference relations," *IEEE T. Cybern.*, vol. 51, no. 1, pp. 283–296, Jan. 2021.

[21] F. J. Quesada, I. Palomares, and L. Martínez, "Managing experts behavior in large-scale consensus reaching processes with uninorm aggregation operators," *Appl. Soft. Comput.*, vol. 35, pp. 873–887, Oct. 2015.

[22] Z.-P. Tian, R.-X. Nie, J.-Q. Wang, and R.-Y. Long, "Adaptive consensus-based model for heterogeneous large-scale group decision-making: Detecting and managing noncooperative behaviors," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 8, pp. 2209–2223, Aug. 2021.

[23] J.-L. Zhou and J.-A. Chen, "A consensus model to manage minority opinions and noncooperative behaviors in large group decision making with probabilistic linguistic term sets," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 6, pp. 1667–1681, Jun. 2021.

[24] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.

[25] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011.

[26] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *J. Am. Med. Inf. Assoc.*, vol. 18, no. 5, pp. 544–551, Sep. 2011.

[27] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, “New avenues in opinion mining and sentiment analysis,” *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 15–21, Mar.–Apr. 2013.

[28] K. Schouten, O. van der Weijde, F. Frasincar, and R. Dekker, “Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data,” *IEEE T. Cybern.*, vol. 48, no. 4, pp. 1263–1275, Apr. 2018.

[29] R. Ren and D. Wu, “An innovative sentiment analysis to measure herd behavior,” *IEEE Trans. Syst. Man Cybern. -Syst.*, vol. 50, no. 10, pp. 3841–3851, Oct. 2020.

[30] S. Wen, H. Wei, Y. Yang, Z. Guo, Z. Zeng, T. Huang, and Y. Chen, “Memristive LSTM network for sentiment analysis,” *IEEE Trans. Syst. Man Cybern. -Syst.*, vol. 51, no. 3, pp. 1794–1804, Mar. 2021.

[31] X. Sun, Y. Gao, R. Sutcliffe, S.-X. Guo, X. Wang, and J. Feng, “Word representation learning based on bidirectional GRUs with drop loss for sentiment classification,” *IEEE Trans. Syst. Man Cybern. -Syst.*, vol. 51, no. 7, pp. 4532–4542, Jul. 2021.

[32] M. Tang, H. Liao, and H. Fujita, “Delegation mechanism-based large-scale group decision making with heterogeneous experts and overlapping communities,” *IEEE Trans. Syst. Man Cybern. -Syst.*, vol. 52, no. 6, pp. 3542–3555, Jun. 2022.

[33] R. E. Bellman and L. A. Zadeh, “Decision-making in a fuzzy environment,” *Manage. Sci.*, vol. 17, no. 4, pp. 141–164, Dec. 1970.

[34] H. Bustince, E. Barrenechea, M. Pagola, J. Fernández, Z. S. Xu, B. Bedregal, J. Montero, H. Hagras, F. Herrera, and B. De Baets, “A historical account of types of fuzzy sets and their relationships,” *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 1, pp. 179–194, Feb. 2016.

[35] E. Herrera-Viedma, I. Palomares, C.-C. Li, F. J. Cabrerizo, Y. C. Dong, F. Chiclana, and F. Herrera, “Revisiting fuzzy and linguistic decision making: Scenarios and challenges for making wiser decisions in a better way,” *IEEE Trans. Syst. Man Cybern. -Syst.*, vol. 51, no. 1, pp. 191–208, Jan. 2021.

[36] F. J. Cabrerizo, R. Al-Hmouz, A. Morfeq, A. S. Balamash, M. A. Martínez, and E. Herrera-Viedma, “Soft consensus measures in group decision making using unbalanced fuzzy linguistic information,” *Soft Comput.*, vol. 21, no. 11, pp. 3037–3050, 2017.

[37] Z. Gong, X. Xu, W. Guo, E. Herrera-Viedma, and F. J. Cabrerizo, “Minimum cost consensus modelling under various linear uncertain-constrained scenarios,” *Inf. Fusion*, vol. 66, pp. 1–17, Feb. 2021.

[38] B. Zhang, Y. C. Dong, X. Feng, and W. Pedrycz, “Maximum fuzzy consensus feedback mechanism with minimum cost and private interest in group decision-making,” *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 9, pp. 2689–2700, Sep. 2021.

[39] Z. Zhang and Z. Li, “Personalized individual semantics-based consistency control and consensus reaching in linguistic group decision making,” *IEEE Trans. Syst. Man Cybern. -Syst.*, vol. 52, no. 9, pp. 5623–5635, Sep. 2022.

[40] Z. Li, Z. Zhang, and W. Yu, “Consensus reaching with consistency control in group decision making with incomplete hesitant fuzzy linguistic preference relations,” *Comput. Ind. Eng.*, vol. 170, Aug. 2022, Art. no. 108311.

[41] A. Pradera, G. Beliakov, H. Bustince, and B. De Baets, “A review of the relationships between implication, negation and aggregation functions from the point of view of material implication,” *Inf. Sci.*, vol. 329, pp. 357–380, Feb. 2016.

[42] F. J. Cabrerizo, R. Ureña, W. Pedrycz, and E. Herrera-Viedma, “Building consensus in group decision making with an allocation of information granularity,” *Fuzzy Sets Syst.*, vol. 255, pp. 115–127, Nov. 2014.

[43] W. Liu, Y. C. Dong, F. Chiclana, F. J. Cabrerizo, and E. Herrera-Viedma, “Group decision-making based on heterogeneous preference relations with self-confidence,” *Fuzzy Optim. Decis. Mak.*, vol. 16, no. 4, pp. 429–447, 2017.

[44] D. Sambrano, J. Masip, and I. Blandón-Gitlin, “How emotions affect judgement and decision making in an interrogation scenario,” *Legal Criminol. Psychol.*, vol. 26, no. 1, pp. 62–82, Feb. 2021.

[45] M. Escadas, M. S. Jalali, and M. Farhangmehr, “Why bad feelings predict good behaviours: The role of positive and negative anticipated emotions on consumer ethical decision making,” *Bus. Ethics*, vol. 28, no. 4, pp. 529–545, Oct. 2019.

[46] S. B. Kotsiantis, “Supervised machine learning: A review of classification techniques,” *Inform.-J. Comput. Inform.*, vol. 31, no. 3, pp. 249–268, 2007.

[47] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, and R. Kumar, “Semeval-2019 task 6: Identifying and categorizing offensive language in social media (OffensVal),” 2019, arXiv:1903.08983v3.

[48] G. Angiani, L. Ferrari, T. Fontanini, P. Fornacciari, E. Iotti, F. Magliani, and S. Manicardi, “A comparison between preprocessing techniques for sentiment analysis in Twitter,” in *Proc. 2nd Int. Workshop Knowl. Discovery WEB (KDWeb)*, 2016.

[49] I. Mozetič, M. Grčar, and J. Smailović, “Multilingual Twitter sentiment classification: The role of human annotators,” *PLoS One*, vol. 11, no. 5, 2016, Art. no. e0155036.

[50] S. Kiritchenko, X. Zhu, and S. Mohammad, “Sentiment analysis of short informal texts,” *J. Artif. Intell. Res.*, vol. 50, pp. 723–762, 2014.

[51] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña-López, and R. Mitkov, “Polarity classification for Spanish tweets using the COST corpus,” *J. Inf. Sci.*, vol. 41, no. 3, pp. 263–272, Jun. 2015.

[52] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[53] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley-Interscience, 1973.

[54] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[55] J. R. Trillo, A. Fernández, and F. Herrera, “HFER: Promoting explainability in fuzzy systems via hierarchical fuzzy exception rules,” in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2020, pp. 1–8.

[56] S. A. Orlovsky, “Decision-making with a fuzzy preference relation,” *Fuzzy Sets Syst.*, vol. 1, no. 3, pp. 155–167, Jul. 1978.

[57] M. J. del Moral, F. Chiclana, J. M. Tapia, and E. Herrera-Viedma, “A comparative study on consensus measures in group decision making,” *Int. J. Intell. Syst.*, vol. 33, no. 8, pp. 1624–1638, Aug. 2018.

[58] S. R. Hait, R. Mesiar, P. Gupta, D. Guha, and D. Chakraborty, “The bonferroni mean-type pre-aggregation operators construction and generalization: Application to edge detection,” *Inf. Fusion*, vol. 80, pp. 226–240, Apr. 2022.

[59] R. R. Yager, “The power average operator,” *IEEE Trans. Syst. Man Cybern. Paart A-Syst. Hum.*, vol. 31, no. 6, pp. 724–731, Nov. 2001.

[60] J. R. Trillo, F. J. Cabrerizo, F. Chiclana, M. Á. Martínez, F. Mata, and E. Herrera-Viedma, “Theorem verification of the quantifier-guided dominance degree with the mean operator for additive preference relations,” *Mathematics*, vol. 10, no. 12, 2022, Art. no. 2035.

[61] I. Palomares, *Large Group Decision Making: Creating Decision Support Approaches at Scale*. Cham, Switzerland: Springer Nature, 2018.

[62] Z.-S. Chen, L.-L. Yang, K.-S. Chin, Y. Yang, W. Pedrycz, J.-P. Chang, L. Martínez, and M. J. Skibniewski, “Sustainable building material selection: An integrated multi-criteria large group decision making framework,” *Appl. Soft. Comput.*, vol. 113, Dec. 2021, Art. no. 107903.

[63] L. Xiao, Z.-S. Chen, X. Zhang, J.-P. Chang, W. Pedrycz, and K.-S. Chin, “Bid evaluation for major construction projects under large-scale group decision-making environment and characterized expertise levels,” *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 1227–1242, 2020.

[64] T. Gai, M. Cao, F. Chiclana, Z. Zhang, Y. C. Dong, E. Herrera-Viedma, and J. Wu, “Consensus-trust driven bidirectional feedback mechanism for improving consensus in social network large-group decision making,” *Group Decis. Negot.*, vol. 32, pp. 45–74, 2023.

José Ramón Trillo received his B.Sc. in mathematics and computer science from the University of Granada, Spain, in 2018, and his M.Sc. in computer science also from the University of Granada in 2019.

He is currently pursuing his Ph.D. degree in computer science from the University of Granada. His research interests include decision analysis, machine learning, consensus modeling, sentiment analysis, computing with words, and large-scale group decision-making.

Enrique Herrera-Viedma (IEEE Fellow) received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Spain, in 1993 and 1996, respectively. He is currently a Professor in Computer Science and Artificial Intelligence and the Vice-President of Research and Knowledge Transfer at the University of Granada. His h-index is 93 with more than 34500 citations received in Web of Science. He has been identified as a Highly Cited Researcher by Clarivate Analytics in both Computer Science and Engineering from 2014 to 2022. He was the 2019-2020 Vice President for Publications with the IEEE System Man & Cybernetics Society and an Associate Editor for several journals, such as IEEE Trans. Fuzzy Syst., IEEE Trans. Syst. Man Cybern. -Syst., IEEE Trans. Intell. Transp. Syst., Inf. Sci., Appl. Soft. Comput., Soft Comput., Fuzzy Optim. Decis. Mak., and Knowledge-Based Syst. His current research interests include group decision-making, consensus models, linguistic modelling, information aggregation, information retrieval, bibliometrics, digital libraries, web quality evaluation, recommender systems, and social media.

Juan Antonio Morente-Molinera received the Ph.D. degree in information and communications technology from the University of Granada, Spain, in 2015.

He is currently an Assistant Professor at the Department of Computer Science and Artificial Intelligence, University of Granada. He is a member of the Andalusian Research Institute of Data Science and Computational Intelligence, DaSCI. He is working in the decision-making field trying to solve the challenges that the new era of ubiquitous computing brings to the decision-support systems field. His research interests include group decision-making, decision support systems, consensus models, linguistic modelling, and information aggregation.

Francisco Javier Cabrerizo received his M.Sc. and Ph.D. degrees in computer science from the University of Granada, Spain, in 2006 and 2008, respectively.

He is currently an Associate Professor with the Department of Computer Science and Artificial Intelligence, University of Granada, Spain. He is an Associate Editor of IEEE T. Cybern., J. Intell. Fuzzy Syst., and Int. J. Inf. Technol. Decis. Mak. His h-index is 34 and he presents more than 5600 citations in Web of Science from Clarivate Analytics. He has been identified as a Highly Cited Researcher by Clarivate Analytics from 2018 to 2022. His research interest includes fuzzy decision-making, decision support systems, consensus models, linguistic modelling, aggregation of information, digital libraries, web quality evaluation, and bibliometrics.