Expert Systems With Applications 263 (2025) 125726

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

# A data-driven large-scale group decision-making framework for managing ratings and text reviews

Diego García-Zamora [a], Bapi Dutta [b,*], LeSheng Jin [c], Zhen-Song Chen [d,e], Luis Martínez [b]

[a] Department of Mathematics, University of Jaén, Jaén, Spain
[b] Department of Computer Science, University of Jaén, Jaén, Spain
[c] Business School, Nanjing Normal University, Nanjing, China
[d] School of Civil Engineering, Wuhan University, Wuhan 430072, China
[e] Faculty of Architecture, Civil Engineering and Applied Arts, Academy of Silesia, 40-555 Katowice, 43 Rolna Street, Poland

ARTICLE INFO

*Keywords:*
STandR-BUI
Basic uncertain information
Large-scale group decision-making
Microblogging
Data-driven decision-making
Consensus

ABSTRACT

Even though the integration of sentiment analysis and decision-making techniques has become popular in recent years, most of the related studies only consider the obtained sentiment score, thus neglecting the numerical ratings that are usually attached to text reviews. This paper introduces STandR (Sentiment from Text and Ratings)-BUI (Basic Uncertain Information), a novel preference-modeling structure for data-driven decision-making using social media microblogging information. STandR-BUI combines both the numerical rating and the sentiment score of a product into a BUI value, which provides a more precise representation of users' opinions. In addition, we propose a consensus framework to make decisions based on the STandR-BUI values which can manage thousands of user reviews. Finally, an illustrative example is provided to demonstrate its effectiveness.

## 1. Introduction

While data-driven models have gained popularity in recent years, expert-driven approaches remain crucial for addressing decision situations where quantitative data is unavailable. In such cases, the only source of information is the qualitative opinions provided by experts in the field (García-Zamora et al., 2022). However, since each individual has a unique background, perspective, and interests, relying solely on one person's preferences to make a decision may not be desirable (Herrera-Viedma et al., 2002). To mitigate the impact of such subjectivity on expert-driven methodologies, group decision-making (GDM) has emerged for complex decision situations in which multiple decision-makers interact to obtain a collective solution for the problem (Song et al., 2024). By leveraging a group's diverse insights and knowledge, GDM seeks to enhance the robustness and inclusivity of the decision-making process, effectively balancing the potential biases and limitations of individual decision-makers (DMs) (Herrera-Viedma et al., 2014).

The digital era and the widespread adoption of novel technologies have led to a shift toward new approaches to solving real-world problems. In the field of decision-making, GDM problems have evolved from involving a few experts to engaging a large number of DMs, leading to large-scale group decision-making (LSGDM) (Zhang et al., 2017). E-democracy technologies (Kim, 2008), e-marketplaces (Karthik & Ganapathy, 2021), social media (Sueur et al., 2012), and disaster management scenarios (Xu et al., 2018) are just a few examples of new LSGDM situations that require the participation of an increasing number of DMs. In this respect, the effectiveness of LSGDM depends not only on the decision methodology, but also on the quality of the data and information available, and the ability to reach collective solutions that satisfy all the members of the group (García-Zamora et al., 2022).

GDM problems were originally conceived to integrate the knowledge of different DMs to mitigate individual subjectivity in decision-making (Labella et al., 2020). With the advent of LSGDM, this concept was expanded to include not just DMs but also the general public, such as city populations or social network users, whose opinions are vital for decisions affecting broader demographics (García-Zamora et al., 2022). This shift necessitates advanced methods to harness users' opinions from online platforms effectively. In this landscape, Sentiment Analysis (SA) has become indispensable for numerically translating users' opinions from the text on platforms like microblogging services, where the brevity of expression is common (Hermida, 2010; Medhat et al., 2014). Recently, researchers have explored various SA methodologies within multi-criteria decision-making (MCDM) frameworks to

* Corresponding author.
*E-mail addresses:* dgzamora@ujaen.es (D. García-Zamora), bdutta@ujaen.es (B. Dutta), jls1980@163.com (L. Jin), zschen@whu.edu.cn (Z.-S. Chen), martin@ujaen.es (L. Martínez).

https://doi.org/10.1016/j.eswa.2024.125726
Received 10 January 2024; Received in revised form 27 October 2024; Accepted 5 November 2024
Available online 15 November 2024

leverage textual reviews (Zuheros et al., 2023). For instance, studies have utilized hesitant, intuitionistic, and interval type-2 fuzzy sets to model sentiment scores and assess products (Heidary Dahooie et al., 2021; Qin et al., 2023; Zhang et al., 2020). Additionally, Liang et al. (2020) employed Probabilistic Linguistic Term Sets to refine customers' SA toward different attributes in web celebrity shops. Moreover, He and Wang (2023) proposed a picture fuzzy set-based SA framework to evaluate new energy vehicles using online reviews. Cao et al. (2022) also integrated topic-mining with SA to improve attribute information acquisition and risk management in emergency decision-making.

SA methods have demonstrated enormous applicability in addressing LSGDM problems involving textual online reviews. However, most existing proposals predominantly focus on the sentiment score derived from text, often neglecting the nuanced insights that numerical ratings can provide, thus oversimplifying users' opinions (Liao et al., 2023). A limited number of studies consider both elements in their analysis. According to Liao et al. (2023), the research that simultaneously considers these two types of reviews is usually devoted to determining the information's consistency and can be categorized as follows:

- Data Analysis: These studies integrate both rating and text reviews for comprehensive data analysis (Chang et al., 2019; Hlee, 2020; Tang et al., 2016). For instance, Tang et al. (2016) found out that positive star ratings and text reviews from bank customers significantly predicted increases in firms' future profitability.
- Correlation Studies: This kind of study explores the correlation between numerical ratings and text reviews (Agarwal et al., 2020; Dong et al., 2020; Shi et al., 2021). For example, Shi et al. (2021) examined the relationships between text reviews, numerical ratings, and the perceived helpfulness of product reviews on Amazon.
- Topic Association: This kind of research investigates how specific topics from text reviews correlate with numerical ratings (Devgan et al., 2020; Jia, 2020). Agarwal et al. (2019) reported that when evaluating urgent care centers, themes like recommendation and prescription refills are usually associated with 5-star ratings, whereas lack of confidence and reception experience is related to 1-star ratings.
- Predictive Modeling: These proposals use the interaction between text and rating reviews to predict additional variables (Cho et al., 2022; Du et al., 2020). For instance, Jerripothula et al. (2020) developed a feature rating system that uses the overall product rating and the text reviews of the features to obtain the ratings of the features.
- Integration into Aggregated Values: This kind of study attempts to combine these reviews into a single aggregated value to improve decision-making (Sharma & Dutta, 2021; Wu & Liao, 2021; Zhu et al., 2022). For instance, Sharma and Dutta (2021) incorporated numerical ratings to enhance SA results.

Despite advancements in SA and decision-making research, several significant challenges persist. Firstly, the integration of ratings and text reviews within decision-making frameworks remains underexplored. Current methodologies are scant and often fail to capture the complexity of user-generated content in online platforms, highlighting the need for innovative approaches that can advance the state-of-the-art (Liao et al., 2023). Secondly, the synthesis of heterogeneous information, such as ratings and text, into a cohesive and manageable format is still an unresolved issue. Effective strategies are required to condense this diverse data into a format that facilitates rapid processing and analysis, particularly when dealing with large volumes of reviews. Moreover, the incorporation of consensus-based approaches in the analysis of online reviews is notably absent in existing literature. Consensus is a cornerstone of traditional LSGDM, and integrating it into online review-based decision-making could transform the aggregation and interpretation of user opinions, leading to a more democratic and representative decision-making process. This would not only enhance the legitimacy of the decisions made but also ensure that they reflect a broader range of user perspectives.

To address these challenges, this paper first introduces the STandR (Sentiment from Text and Ratings)-BUI (Basic Uncertain Information) model, a novel preference-modeling structure designed to encapsulate both the numerical ratings and text reviews into a unified BUI value, which consists of a tuple $(x, c)$, where $x$ is a value within the interval $[0, 1]$ and $c \in [0, 1]$ stands for the degree of certainty regarding the value $x$ (Mesiar et al., 2018). Integrating these different types of data into the BUI value, allows for a more convenient and nuanced representation of user opinions, facilitating more informed decision-making based on data from microblogging platforms.

Building on this foundation, we propose an automated consensus framework designed specifically for data-driven decision-making that leverages the synthesized BUI values. This framework is particularly well-suited for the dynamic and often fragmented environment of social media microblogging, where user opinions are diverse and rapidly evolving. To illustrate the practicality and effectiveness of the STandR-BUI model and the associated consensus framework, we include a detailed example that demonstrates how our model processes real social media data. This example highlights our framework's capability to enhance decision-making and ensure representativeness in real-world scenarios. To summarize, the main novelties of this contribution are as follows:

- A robust decision framework is developed to systematically leverage information extracted from social media user's reviews. This framework is designed to facilitate informed decision-making by analyzing user-generated content, which often embodies diverse insights into user preferences and experiences.
- We introduce STandR-BUI, a novel preference representation format that combines both ratings and text reviews into a BUI value. This approach efficiently captures and summarizes the inherent uncertainties within user-generated content, providing a more reliable basis for decision-making.
- An innovative automatic consensus model for BUI values is proposed, capable of handling and synthesizing opinions from thousands of user reviews. This model obtains collective opinions regarding the performance of alternatives across several criteria, ensuring that the final decisions reflect a wider consensus and enhancing the legitimacy and acceptability of decisions in LSGDM scenarios.

The remainder of this paper is set up as follows. Section 2 introduces the main ideas necessary to understand the proposal. Section 3 develops the theoretical considerations that support our framework, including a method to compute a sentiment score from sentiment probabilities, the definition of STandR-BUI values to model the information within rating and text reviews, and a BUI-based automatic consensus model. In Section 4, we detail our framework for data-driven decision-making. Section 5 develops an illustrative example and Section 6 carries out an analysis of the performance of our method. Finally, Section 7 draws some conclusions.

## 2. Preliminaries

This section introduces the basic concepts necessary to understand our proposal.

### 2.1. Large-Scale Group Decision-Making

Traditionally, a Group Decision-Making (GDM) problem has been defined as a decision situation in which several (generally few) DMs are asked to find a common solution (Herrera-Viedma et al., 2014). Typically, a GDM problem consists of a finite set of alternatives, $H = \{H_1, H_2, \ldots, H_K\}$ $(K \geq 2)$, representing the potential solutions to be

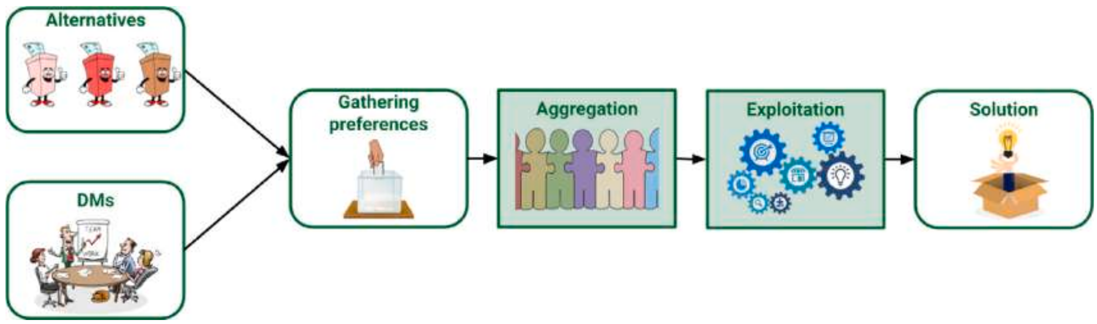



**Fig. 1.** Scheme including the phases of a GDM problem.

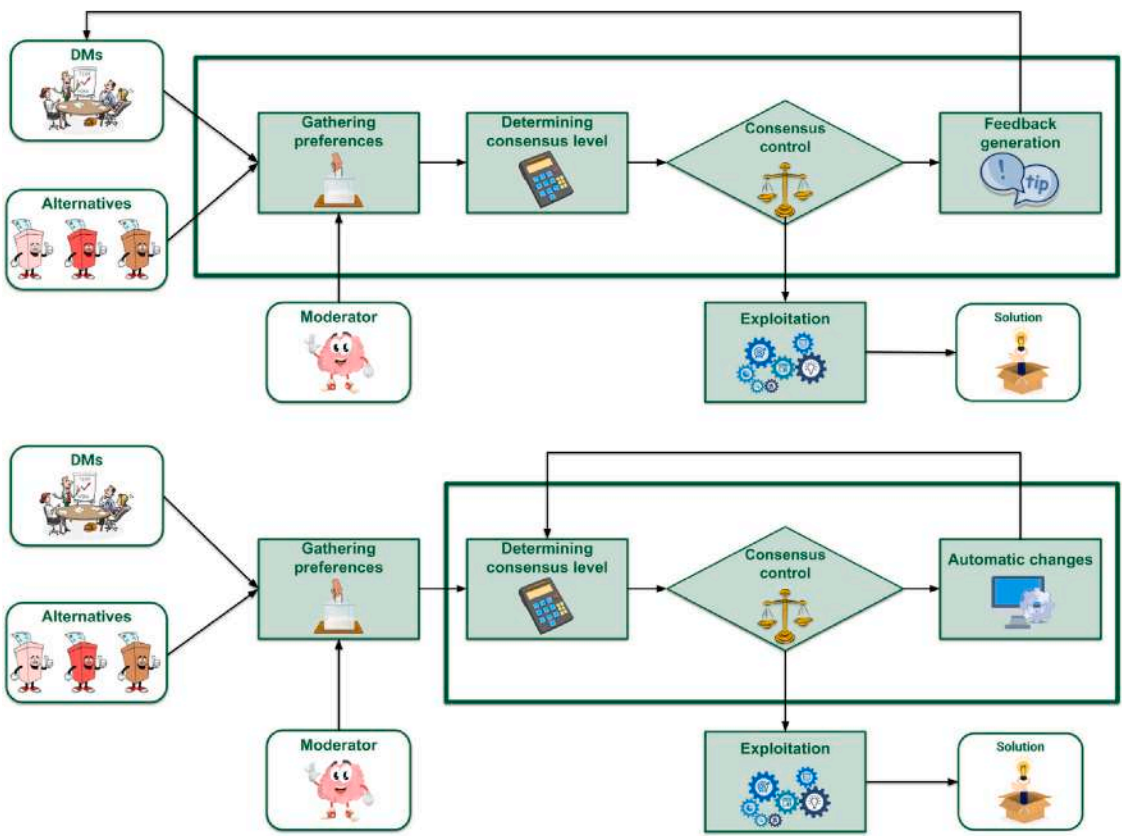



**Fig. 2.** Scheme of feedback-based and automatic CRPs.

evaluated for a given decision-making problem and a group of experts, $E = \{DM_1, DM_2, \ldots, DM_m\}$, where $m \geq 2$, who assess and evaluate these alternatives.

There are several classic rules to model such a decision process, such as the majority and minority rules, or the Borda count (Butler & Rothstein, 1987). Formally, GDM problems involve two main phases, including the aggregation of DMs' opinions into a single collective preference and the exploitation of such a collective preference to select the best alternative (see Fig. 1). Nevertheless, when several DMs collaborate to make a decision, some of them may feel unsatisfied with the decision made by the group, leading to disagreements and even questioning the decision-making process (Herrera-Viedma et al., 2002).

Therefore, in GDM processes, it is usual to integrate a Consensus Reaching Process (CRP) to increase the level of agreement among the DMs in the group (Ben-Arieh & Easton, 2007). Such CRPs are usually iterative discussion processes coordinated by a moderator in which the DMs' original preferences are modified until they are close enough according to the value of a certain consensus measure (Kacprzyk & Fedrizzi, 1988). The general process of a CRP is illustrated in Fig. 2 and can be summarized as follows:

- Gathering Preferences: In this phase, the preferences of each decision-maker (DM) are elicited and collected.
- Consensus Measurement: The group consensus is calculated based on DMs' preferences using specific consensus measures, which are classified into two types based on calculation methods (Palomares et al., 2014):
  - *Distance from Collective Opinion:* Measures consensus by evaluating the distance between each DM's preference and the collective opinion.
  - *Inter-DM Distance:* Measures consensus by evaluating the distance between the preferences of any two DMs in the group.
- Consensus Control: If the group's consensus level is below a predefined threshold, $\mu_0 \in [0, 1]$, additional discussion rounds are required; otherwise, the selection process proceeds.


**Fig. 3.** Scheme including the phases of an LSGDM problem.

- Consensus Progress: A procedure, with or without feedback, is implemented to improve the agreement level across subsequent discussion rounds.

Technological advances have eased the development of decision problems involving hundreds or even thousands of DMs, leading to LSGDM in which, according to the traditional definition, at least twenty DMs are involved (Carvalho et al., 2011). LSGDM techniques allow the integration of the knowledge of many DMs to derive intelligent decisions that account for all their opinions. In this regard, LSGDM shares common phases with GDM, but in the large-scale case, the aggregation phase is much more complex and may include additional steps such as clustering DMs, managing non-cooperative behaviors or social network analysis (see Fig. 3).

In the classic LSGDM literature, CRPs have always been a recurrent topic (Herrera-Viedma et al., 2014). However, it has been pointed out that, in decision problems involving many DMs, the traditional idea of consensus as an iterative process may be inadequate, and the classic feedback mechanisms should be substituted by automatic algorithms that automatically compute a solution that satisfies the involved DMs as much as possible (García-Zamora et al., 2022).

In this context, the Minimum Cost Models (MCC) have emerged as automatic CRPs that reformulate the decision problem in terms of an optimization model consisting of minimizing the cost of modifying DMs' opinions to reach an agreed collective solution (Ben-Arieh & Easton, 2007). These models aim at maximizing a linear similarity/satisfaction between the initial and modified opinions, which makes them especially suitable to manage LSGDM scenarios in a reasonable time (García-Zamora, Dutta, Massanet et al., 2023). While MCC models adjust opinions to align closely with collective views by minimizing a cost function, they overlook traditional consensus measures commonly used in CRP literature (Garcí a-Zamora, Dutta, Labella et al., 2023). Thus, MCC models cannot ensure a desired agreement level. To address this, Rodríguez et al. (2021) introduced CMCC models, as defined mathematically below:

$$\min \quad \sum_{k=1}^{m} c_k |\bar{o}_k - o_k|$$
$$\text{s.t.} \quad \begin{cases} \bar{g} = \sum_{k=1}^{m} w_k \bar{o}_k \\ |\bar{o}_k - \bar{g}| \le \varepsilon, \ k = 1, 2, \ldots, m \\ Consenus(\bar{o}_1, \bar{o}_2, \ldots, \bar{o}_m) \ge \mu_0 \end{cases} \tag{CMCC}$$

where $(\bar{o}_1, \ldots, \bar{o}_m)$ represent the adjusted opinions of the experts for initial opinions $(o_1, \ldots o_k)$, $\bar{g}$ denotes the collective opinion, and $\varepsilon$ is the maximum allowable distance between each expert's opinion and the collective opinion, $\mu_0$ denotes the consensus threshold and $(c_1, \ldots, c_m)$ is the costs of adjusting opinions of the experts.

### 2.2. Basic uncertain information

The Basic Uncertain Information (BUI) model, introduced by Mesiar et al. (2018), offers an alternative preference structure for managing uncertainty in decision-making scenarios. A BUI value is represented as a pair $\langle x; c\rangle$, where $x \in [0,1]$ denotes the degree of preference for an alternative, and $c \in [0,1]$ expresses the certainty that the value is accurate. In the context of expert evaluations, $x$ reflects the expert's preference or opinion, while $c$ represents the reliability of the expert's judgment. An illustrative example of BUI values can be seen in the peer review process for journal or conference submissions. Using a linguistic scale instead of an interval scale, a BUI value such as $\langle$weak accept; medium$\rangle$ might indicate that a referee recommends a weak acceptance for a paper and self-assesses their expertise on the topic as a medium.

In the decision-making process, one of the key aspects is the aggregation of information. To aggregate these BUI values, Mesiar et al. (2018) proposed the following generic aggregation mechanism:

**Definition 1.** Let $J = \{\langle x; c\rangle : x, c \in [0,1]\}$ and consider the projections $P_x : J \to [0,1]$ and $P_c : J \to [0,1]$ defined as $P_x(\langle x; c\rangle) = x$, $P_c(\langle x; c\rangle) = c \ \forall \langle x; c\rangle \in J$, respectively. A mapping $\tilde{A} : J^n \to J$ is said to be a BUI- aggregation function if there exists a mapping $A : [0,1]^n \to [0,1]$ such that

1. $A(x_1, \ldots, x_n) = P_x(\tilde{A}(\langle x_1; c_1\rangle, \ldots, \langle x_n; c_n\rangle)) \ \forall \langle x_1; c_1\rangle, \ldots, \langle x_n; c_n\rangle \in J$.
2. For any $x \in [0,1]^n$ the mapping $A_x : [0,1]^n \to [0,1]$ given by
$$A_x(c_1, \ldots, c_n) = P_c(\tilde{A}(\langle x_1; c_1\rangle, \ldots, \langle x_n; c_n\rangle))$$
is an aggregation function.
3. $A_x(c) = A_y(c) \ \forall x, y, c \in [0,1]$ such that $c_i x_i = c_i y_i \ \forall i = 1, 2, \ldots, n$;

Another important result in BUI aggregation is the transformation of the BUI value to a unique interval through bijective mappings, which allows aggregation of BUI values through interval aggregation. These bijective mappings are as follows (Mesiar et al., 2018):

- $\phi : J^* \to L([0,1])^*$ defined as $\phi(\langle x; c\rangle) = [cx, 1 - c + cx] \ \forall \langle x; c\rangle \in J^*$,
- $\phi^{-1} : L([0,1])^* \to J^*$ defined as $\phi^{-1}([a,b]) = \langle \frac{a}{1-b+a}; 1 - b + a\rangle \ \forall [a,b] \in L([0,1])^*$,

where $J^* := \{\langle x; c\rangle : x \in [0,1], c \in ]0,1]\}$ and
$L([0,1])^* := \{[a,b] : a, b \in [0,1], b - a < 1\}$. These mappings guarantee that any BUI whose certainty degree is non-null can be univocally remapped into an interval strictly contained in $[0,1]$ and vice versa.

## 3. Some technical considerations

The ultimate purpose of this study is to develop an automatic decision model that accounts for the heterogeneous opinions (ratings and text reviews) provided by users on social media regarding the alternatives/products over different criteria/features. As in any other LSGDM model, such a framework requires modeling users' opinions, providing an aggregation methodology to combine these opinions into a collective representation, and finally giving a method to determine, from the collective preference, which alternative is the best one.

This section is devoted to developing the theoretical basis that supports our decision framework. Specifically, we first address the modeling of users' opinions, focusing on how to mathematically define a sentiment score for a text based on the probabilities of it being positive, neutral, or negative. After clarifying the computation of the sentiment score, we present a unification methodology to represent a user review consisting of both a rating and a text as a single BUI value. This will result in a novel preference structure called STandR-BUI. Afterward, we define an automatic consensus model based on MCC that processes STandR-BUI values as inputs and produces a single STandR-BUI value representing the collective opinion of all users. Theoretically, this aggregated value reflects the collective opinion by adjusting individual users' opinions with the minimum necessary changes to achieve a specified consensus degree.

### 3.1. Computing a sentiment score from probabilities

SA algorithms often provide the sentiment of a text utilizing probability values (Pang et al., 2002; Tang et al., 2014). Formally, given the linguistic term set (LTS) $\{s_-, s_., s_+\}$, where

$s_- =$ {Negative sentiment},

$s_. =$ {Neutral sentiment},

$s_+ =$ {Positive sentiment},

the sentiment of a microblogging text may be expressed in terms of probability-based linguistic term sets as follows:

$$\overline{S} = \{(s_-, p_-), (s_., p_.), (s_+, p_+) \,:\, p_-, p_., p_+ \geq 0, p_- + p_. + p_+ = 1\},$$

where $p_-, p_., p_+$ are the probabilities that the sentiment is negative, neutral, or positive. In the rest of the manuscript, we refer to the aforementioned set $\overline{S}$ as the *Sentiment Term Set* (STS). Note that there exists an obvious bijection between the STS and the plane section

$$S = \{(x, y, z) \in [0, 1]^3 \,:\, x + y + z = 1\} = \{(p_-, p_., p_+) \,:\, p_-, p_., p_+ \geq 0,$$
$$p_- + p_. + p_+ = 1\},$$

where the coordinates $x, y, z$ represent, respectively, the probability of negativity, neutrality, or positivity of the text. The following result provides an additional bijection to represent the probabilities in the plane section $S$ as an element of the set $L([0, 1]) := \{[a, b] \subseteq [0, 1] \,:\, a \leq b\}$.

**Proposition 1.** *The mapping* $F \,:\, S \to L([0, 1])$ *defined as:*

$$F(x, y, z) = [z, 1 - x] \;\forall\; (x, y, z) \in S$$

*is a bijection whose inverse:* $F^{-1} \,:\, L([0, 1]) \to S$ *is given by*

$$F^{-1}([a, b]) = (1 - b, b - a, a) \;\forall\; [a, b] \in L([0, 1]).$$

The mapping $F$ allows remapping an evaluation of the sentiment of a text expressed through the probabilities $(s_-, p_-), (s_., p_.), (s_+, p_+)$ as the interval $[p_+, 1 - p_-]$ contained in $[0, 1]$. In addition, $F$ presents the following properties:

- The closer the interval is to $[0, 0]$, the more negative the sentiment of the text, i.e., $F(1, 0, 0) = [0, 0]$.
- The closer the interval is to $[1, 1]$, the more positive the sentiment of the text, i.e., $F(0, 0, 1) = [1, 1]$.
- Even though the interval just depends on the probabilities corresponding to the positive and negative sentiment, the information regarding the neutrality is codified as the length of the interval $[p_+, 1 - p_-]$. In this regard, the wider the interval, the more neutral the sentiment of the text, i.e., $F(0, 1, 0) = [0, 1]$.

Although the mapping $F$ provides a representation of the sentiment of a text as an interval, some situations require the utilization of a single crisp value that summarizes the sentiment of the text (Zuheros et al., 2023). To construct such a score function, here we introduce a method based on proportionality. First, let us consider the functions $\mu_-, \mu_., \mu_+ \,:\, [-1, 1] \to [0, 1]$ defined as:

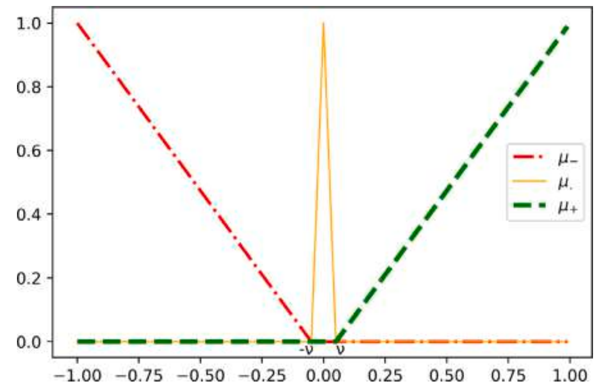



**Fig. 4.** Graphs of the functions for sentiment labels $\mu_-, \mu_.$, and $\mu_+$.

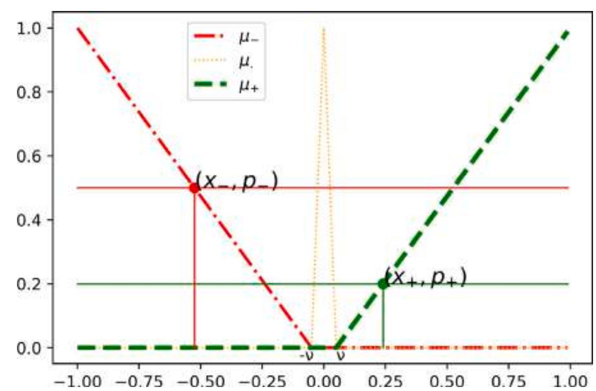



**Fig. 5.** Projections $x_-, x_+$ used to transform the sentiment of a text given in terms of probabilities $(p_-, p_., p_+)$ into a numerical score.

$$\mu_-(x) = \begin{cases} 1 - \frac{x+1}{1-v} & \text{if } -1 \leq x \leq -v \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_.(x) = \begin{cases} \frac{x+v}{v} & \text{if } -v \leq x \leq 0 \\ 1 - \frac{x}{v} & \text{if } 0 < x \leq v \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_+(x) = \begin{cases} \frac{x-v}{1-v} & \text{if } v \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

where $v \in\, ]0, 1[$ is a parameter for controlling the shape of the functions corresponding to the sentiment labels (see Fig. 4). The parameter $v$ provides some flexibility when determining the sentiment score according to the decision problem. Traditionally, a text is said to be neutral if its score value expressed in a bipolar unit scale is between $[-0.05, 0.05]$.[1] This suggests that the length of the support of the function for "neutrality" should be lower than the length of the support of the functions for "negativity" or "positivity". Nevertheless, other authors may have a different vision and consider other thresholds to define the boundary for negative or positive sentiment (Borg & Boldt, 2020). In this regard, we provide a sensitivity analysis in Section 5 to give some insights into the impact of this parameter.

Given a text whose sentiment is described by $(p_-, p_., p_+) \in S$, it is necessary to obtain a value in the interval $[0, 1]$ that represents the aggregated value/score of these probabilities. To transform these probabilities into values on the bipolar scale $[-1, 1]$, we first obtain the values $x_- \in [-1, -v], x_+ \in [v, 1]$ such that $\mu_-(x_-) = p_-$ and $\mu_+(x_+) = p_+$ (see Fig. 5). Concretely, these values are given as:

[1] https://communalytic.com/video-tutorials/tutorial-sentiment-analysis/.

$$x_- = (1-\nu)(1-p_-) - 1,$$
$$x_+ = \nu + (1-\nu)p_+.$$

Then, to obtain an overall value that represents the sentiment of a text on the bipolar scale $[-1,1]$, we aggregate the values $x_-$ and $x_+$ using a weighted average whose weights are the probabilities $p_-$ and $p_+$:

$$\begin{aligned} x_-p_- + x_+p_+ &= ((1-\nu)(1-p_-) - 1)p_- + (\nu + (1-\nu)p_+)p_+ \\ &= \nu(p_+ - p_-) + (1-\nu)(p_+^2 - p_-^2) \in [-1,1]. \end{aligned}$$

Finally, we can define a score function $Sc : S \to [0,1]$ as follows:

$$Sc(p_-, p_\cdot, p_+) := \frac{\nu(p_+ - p_-) + (1-\nu)(p_+^2 - p_-^2) + 1}{2} \tag{1}$$

$\forall\ (p_-, p_\cdot, p_+) \in S$. Note that this function does not directly incorporate the value of $p_\cdot$. However, this scoring mechanism is sensible because:

- The function $F : S \to L([0,1])$ ensures that the probability $p_\cdot$ is not necessary to describe the information in the STS $(p_-, p_\cdot, p_+)$.
- If one follows Eq. (1) to compute the values $x_\cdot$ and $x'_\cdot$ which represent the projections in the abscissa of the intersections of $\mu_\cdot$ and the line $y = p_\cdot$, the aggregation of $x_\cdot$ and $x'_\cdot$ under the same weight is equal to 0.

Additionally, it should be pointed out that, in the case of having chosen an SA algorithm that allows directly obtaining a sentiment score valued on an interval scale, it is possible to go ahead with such a score if it is properly normalized (Borg & Boldt, 2020; Hutto & Gilbert, 2014).

### 3.2. STandR-BUI values to unify rating and text reviews

In this section, we propose a process to automatically model DM's preference in microblogging reviews, using both textual content and numerical ratings. This process relies on the sentiment score function defined in the previous section and is designed for application in real-world decision-making scenarios. Specifically, the ratings provided by the users are combined with the score obtained from the SA of their corresponding text review.

Note that we can formulate this situation from the classical GDM point of view. In this case, the DMs stand for the users, whereas the opinions are their preferences collected from microblogging sites. Therefore, let us assume that in such a GDM problem, each user $U_k$, $k = 1,\dots,K$ expresses their opinion on alternative $H_i$, $i = 1,\dots,m$ under criterion $C_j$, $j = 1,\dots,n$ using both a rating $r_{ij}^k \in [0,1]$ and a text review $t_{ij}^k$. From the text review $t_{ij}^k$, we can apply an SA algorithm to obtain an STS $(p_-, p_\cdot, p_+)_{ij}^k \in S$. We then use the score function $Sc : S \to [0,1]$ defined previously to calculate a sentiment score $s_k^{ij} := Sc((p_-, p_\cdot, p_+)_{ij}^k)$, which summarizes the sentiment of the text review $t_{ij}^k$ on a scale $[0,1]$.

Note that both $r_k^{ij}$ and $s_k^{ij}$ are expressed on the same scale and represent two evaluations provided by the same user on the same alternative and according to the same criteria. However, the values $t_k^{ij}$ and $s_k^{ij}$ do not necessarily have to be equal. The greater the difference between these values, the more inconsistent user $U_k$'s opinion is considered to be. In other words, when the distance between $t_k^{ij}$ and $s_k^{ij}$ is high, the certainty about the user $U_k$'s real opinion on the alternative $H_i$ according to $C_j$ should be low. In this context, such uncertainty can be conveniently modeled by BUIs.

**Definition 2** (*STandR-BUI Preference*)**.** Let $r_{ij}^k \in [0,1]$ and $s_{ij}^k := Sc((p_-, p_\cdot, p_+)_{ij}^k) \in [0,1]$ be, respectively, the rating and the sentiment score of the text provided by the user $U_k$ on the alternative $H_i$ under the criterion $C_j$. The corresponding STandR (Sentiment from Text and Rating)-BUI preference is defined as

$$\langle x_{ij}^k; c_{ij}^k \rangle := \begin{cases} \phi^{-1}([\min\left\{r_{ij}^k, s_{ij}^k\right\}, \max\left\{r_{ij}^k, s_{ij}^k\right\}]) & \text{if } |r_{ij}^k - s_{ij}^k| < 1 \\ \langle \frac{1}{2}; 0 \rangle & \text{otherwise} \end{cases}. \tag{2}$$

Note that the value $\frac{1}{2}$ in $\langle \frac{1}{2}; 0 \rangle$ is selected ad hoc since the certainty coefficient is null. Nevertheless, the emergence, in practice, of a situation in which $|r_{ij}^k - s_{ij}^k| = 1$ implies that the information provided by the user is completely contradictory, and it should be considered if such an opinion should be neglected.

We further observe that STandR-BUI preferences allow combining the sentiment score $s$ obtained from textual reviews and the numerical rating $r$ provided by users into a single BUI value $\langle x, c \rangle$. Here, $x$ represents an aggregated value of both $r$ and $s$, which can be considered a normalized score. Additionally, $c$ models the certainty associated with the user's evaluation. A low $c$ value indicates that $r$ and $s$ are significantly different, suggesting that $x$ may not be very reliable. Conversely, a $c$ value close to 1 indicates that $r$ and $s$ are consistent and thus $x$ is a reliable representation of both values.

### 3.3. An automatic consensus model for BUI values

In this section, we describe an MCC-based automatic consensus model to obtain consensual collective opinions from the initial users' preferences modeled as STandR-BUI values. Let us assume that the user $U_k$'s opinion ($k = 1,\dots,K_i$) on the alternative $H_i, i = 1,2,\dots,m$ under the criterion $C_j, j = 1,2,\dots,n$ is represented as the STandR-BUI value $\langle x_{ij}^k; c_{ij}^k \rangle$, as previously defined. We now introduce the MCC model to achieve consensus on the collective preference for an alternative $H_i$.

$$\min_{x'^k_{ij} \in [0,1]} \sum_{k=1}^{K_i} \sum_{j=1}^{n} \xi_{ij}^k |x'^k_{ij} - x_{ij}^k|$$
$$s.t. \begin{cases} x'_{ij} = \sum_{k=1}^{K_i} w_k x'^k_{ij}, j = 1,\dots,n \\ |x'^k_{ij} - x'_{ij}| \le \varepsilon, k = 1,\dots,K_i, j = 1,\dots,n \\ 1 - \sum_{k=1}^{K_i} \sum_{j=1}^{n} w_k \omega_j |x'^k_{ij} - x'_{ij}| \ge \mu_0 \end{cases} \tag{3}$$

where $K_i$ represents the number of reviews related to the alternative $H_i$, $w = (w_1,\dots,w_{K_i})$ are the weights assigned to the users, and $\omega = (\omega_1,\dots,\omega_n)$ are the weights for the criteria/features. The consensus threshold is denoted by $\mu_0 \in [0,1]$, and $\epsilon > 0$ measures the maximum allowed distance between individual user opinions and the collective opinion. The relative costs $\xi_{ij}^k > 0$ satisfy $\sum_{k=1}^{K_i} \sum_{j=1}^{n} \xi_{ij}^k = 1$.

From the above optimization model, it is possible to derive modified preferences $x'^k_{ij}$ which are as similar as possible to the original values in the STandR-BUIs $\langle x_{ij}^k; c_{ij}^k \rangle$ and satisfy the consensus conditions. However, the degrees of certainty $c'^k_{ij}$ corresponding to the data $x'^k_{ij}$ are not determined yet. Before defining a method for computing them, let us provide a fuzzy interpretation of the aforementioned bijection $\phi : J^* \to L([0,1])^*$, which forms the basis of computing modified certainty degree.

For $\langle x; c \rangle \in J^*$ such that $x \in ]0,1[$ let us consider the membership function $\mu_x : [0,1] \to [0,1]$ defined as

$$\mu_x(t) = \begin{cases} \frac{t}{x} & \text{if } 0 \le t < x, \\ 1 - \frac{t-x}{1-x} & \text{if } x \le t \le 1. \end{cases}$$

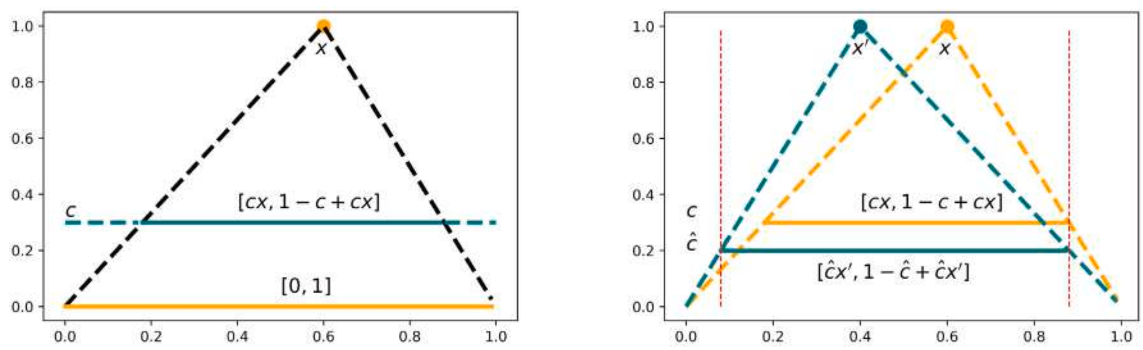
Note that for any $c \in ]0,1[$, the following equality holds:

$$\phi(\langle x; c \rangle) = [cx, 1 - c + cx] = \left\{t \in [0,1] \ : \ \mu_x(t) \ge c\right\} = \mu_x^c$$

where $\mu_x^c$ denotes the classical $\alpha$−cut for $\alpha = c$ of the fuzzy set $\mu_x$. In other words, the mapping $\phi$ transforms the BUI $\langle x; c \rangle$ into the $c$−cut of $\mu_x$ (see Fig. 6(a)).

Let us recall now the problem of determining the modified uncertainty degree $c'^k_{ij}$ for given BUI $\langle x; c \rangle \in J^*$ and a modified data $x'$. Given the aforementioned fuzzy interpretation, the lower feasible value for $c'$ should be the maximum $\alpha \in [0,1]$ such that the $\alpha$-cut of $\mu_{x'}$ contains the $c$-cut of $\mu_x$ (see Fig. 6(b)). Formally:

$$\begin{aligned} c' \ge \hat{c} &:= \max\left\{\alpha \in [0,1] \ : \ \mu_x^c \subseteq \mu_{x'}^\alpha\right\} \\ &= \max\left\{\alpha \in [0,1] \ : \ [cx, 1 - c + cx] \subseteq [\alpha x', 1 - \alpha + \alpha x']\right\}. \end{aligned}$$




(a) Fuzzy interpretation of the bijection $\phi$.

(b) Definition of $\hat{c}$.

**Fig. 6.** Graphs to visualize the selection of $\hat{c}$.

Some algebraic computations on the above expression lead to the following equality:

$$\hat{c} = \begin{cases} c \cdot (1-x) & \text{if } x' = 0 \\ c \cdot x & \text{if } x' = 1 \\ c \cdot \min\left\{ \frac{x}{x'}, \frac{1-x}{1-x'} \right\} & \text{if } x' \in ]0,1[. \end{cases}$$

Consequently, we define $c'$ as a convex combination of both $c$ and $\hat{c}$:

$$c' := \hat{c}|x - x'| + c(1 - |x - x'|). \tag{4}$$

Note that in the case $c = 0$, it must be $c' = 0$. In addition, this procedure to determine $c'$ ensures that if $x' \neq x$, then $c' < c$. At this stage, for each STandR-BUI value $\langle x_{ij}^k; c_{ij}^k \rangle$, we have obtained the modified individual preference $\langle {x'}_{ij}^k; {c'}_{ij}^k \rangle$. Additionally, we have computed the collective value ${x'}_{ij}$. To complete the CRP, we still need to define how to obtain the collective certainty degree ${c'}_{ij}$. To do so, we adopt the standard aggregation functions for BUIs proposed by Mesiar et al. (2018). Therefore, let us consider the BUI aggregation function $A = (A_1, A_2) : J^n \to J$ such that $A_1$ is the weighted average operator and the certainty degree is computed as follows:

$$\begin{aligned} {c'}_{ij} &= A_2(\langle {x'}_{ij}^1; {c'}_{ij}^1 \rangle, \ldots, \langle {x'}_{ij}^K; {c'}_{ij}^K \rangle) \\ &= 1 - \mathcal{L}(\mathcal{B}(\phi(\langle {x'}_{ij}^1; {c'}_{ij}^1 \rangle), \ldots, \phi(\langle {x'}_{ij}^K; {c'}_{ij}^K \rangle))), \end{aligned} \tag{5}$$

where $\mathcal{L} : L([0,1]) \to [0,1]$ is the length operator defined as $\mathcal{L}([a,b]) = b - a, \forall\ [a,b] \in L([0,1])$ and $\mathcal{B} : L([0,1])^K \to L([0,1])$ is the interval aggregation given by

$$\mathcal{B}([a_1, b_1], \ldots, [a_K, b_K]) = [M(a_1, \ldots, a_K), M(b_1, \ldots, b_K)]$$

where $M : [0,1]^K \to [0,1]$ is arithmetic mean operator. Note that the aggregation provided by $A_2$ defined in this way satisfies the certainty aggregation axioms provided by Jin et al. (2018).

## 4. A decision-making framework to integrate ratings and text reviews

This section describes in detail our framework to make decisions based on the information given by social media users through ratings and text reviews. Such a framework is based on an automatic consensus model designed to aggregate diverse opinions from heterogeneous online reviews, enabling users to capture a broad spectrum of user experiences. This model is pivotal in synthesizing a collective opinion that reflects the sentiments of both majority and minority user groups effectively.

First, we segregate each user's textual opinions according to the different criteria, aspects, or features for which ratings have been provided. From the segregated texts of each criterion, we extract the user's sentiment regarding each criterion in terms of probability. Afterward, we use the intuitive score function constructed in the previous section to map sentiment probability into the same scale as the rating. Subsequently, we model the text rating and sentiment score into BUI based on the fact that the subtle difference between the rating and sentiment score would impact the reliability of the rating. Then, the above-defined MCC-based CRP is applied to fuse the BUI information in the multi-criteria decision-making context, leading to the determination of a consensus solution. A high-level scheme for managing the decision-making process with this framework is illustrated in Fig. 7.

### 4.1. Problem description

To align with the earlier notations concerning the decision elements, we consider that the set of alternatives to be assessed is denoted as $H = \{H_1, H_2, \ldots, H_m\}$, and establish the criteria for their evaluation as $C = \{C_1, C_2, \ldots, C_n\}$. Additionally, let $\mathcal{U}_i = \{U_1^i, \ldots, U_{K_i}^i\}$ represent the set of $K_i$ users who evaluate each alternative $H_i$ based on these criteria. These evaluations include both numerical ratings and textual reviews. We further assume that $T_i^k$ represents the textual review submitted by the user $U_k^i$ for alternative $H_i$.

It should be highlighted that the number of reviews available for two different alternatives is not necessarily the same. Indeed, the users providing reviews for different alternatives do not need to be the same individuals.

### 4.2. Text segregation

In the process of textual reviewing, users typically provide detailed accounts of their experiences across various criteria. In real-world scenarios, it is common for users to focus on certain criteria (aspects) while possibly omitting others. Effectively discerning users' opinions across different criteria is crucial for making accurate decisions. To achieve this, we employ Stanford CoreNLP,[2] a Natural Language Processing (NLP) library capable of detecting criteria terms based on part-of-speech tagging and sentence dependency structures, guided by a predefined set of rules (Manning et al., 2014; You et al., 2022). Using this criteria detection tool, we partition the textual review $T_i^k$ provided by the $k$th user for alternative $H_i$ into $n$ segments $T_i^k = (t_{i1}^k, \ldots, t_{in}^k)$ corresponding to the criteria. Furthermore, a user may elaborate on a criterion using multiple sentences, denoted by $t_{ij}^k = \{\tau_{ij1}^k, \ldots, \tau_{ijN_{ij}^k}^k\}$, where $N_{ij}^k$ represents the cardinality of sentences related to criterion $C_j$ for all $k = 1, 2, \ldots, K_i$ and $j = 1, 2, \ldots, n$. This approach allows us to extract the user's feedback on each criterion from their textual review.

[2] https://stanfordnlp.github.io/CoreNLP/.

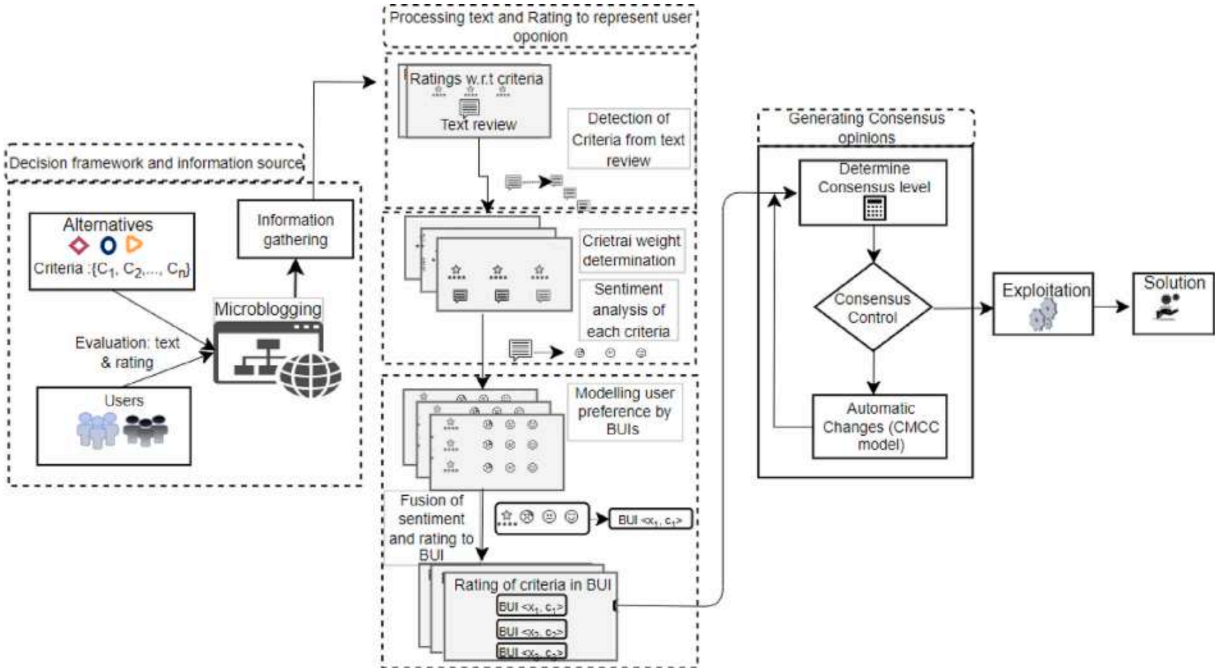



**Fig. 7.** The scheme of the proposed framework for LSGDM with text and rating reviews.

### 4.3. Sentiment analysis

Since our primary goal is to integrate both the numerical rating and the textual review for evaluation, we aim to convert the text into a numeric representation. A practical approach to achieve this is by leveraging the sentiment expressed in the textual review and then incorporating it into the STandR-BUI preference framework to combine it with the rating. Thus, we aim to derive the sentiment probabilities $(p_{\_}p_{.}, p_{+})_{ij}^{k}$ of the texts $t_{ij}^{k}$ against a criterion $C_j$ for an alternative $H_i$ by the $k$th user. There exist various methods to obtain the sentiment of the texts, and here we have adopted the simple lexicon-based approach implemented in `TextBlob`.[3]

In the aim of obtaining sentiment probability of $t_{ij}^{k}$, the sentiment of each $\tau_{ijl}^{k} \in t_{ij}^{k}$ is cast based on the `TextBlob` sentiment analyzer, represented via the linguistic terms $s_{\_}$, $s_{.}$, or $s_{+}$ and denoted as $Senti(\tau_{ijl}^{k})$. Let $N_{ij+}^{k} = |\{\tau_{ijl}^{k} \in t_{ij}^{k} : Senti(\tau_{ijl}^{k}) = s_{+}\}|$ be the total count of the texts with $s_{+}$ and similarly we define $N_{ij-}^{k}$ and $N_{ij.}^{k}$ as counts for $s_{\_}$ and $s_{.}$. Based on the frequency definition of probability, the sentiment probability for $t_{ij}^{k}$ is set as $(p_{\_}, p_{.}, p_{+})_{ij}^{k} = (\frac{N_{ij-}^{k}}{N_{ij}^{k}}, \frac{N_{ij.}^{k}}{N_{ij}^{k}}, \frac{N_{ij+}^{k}}{N_{ij}^{k}})$. From the sentiment probability $(p_{\_}, p_{.}, p_{+})_{ij}^{k}$ of $t_{ij}^{k}$, we obtain the sentiment score $s_{ij}^{k} = Sc((p_{\_}, p_{.}, p_{+})_{ij}^{k})$ by utilizing the Eq. (1). In this manner, we obtain the sentiment score for all the textual reviews against different criteria given by the users.

### 4.4. Criteria weights

The determination of the weights for the criteria is one of the key issues in the decision process. Here, we propose an objective weight determination method based on user textual reviews. Specifically, the weights for the criteria are assigned according to the users' concerns described in the text. If users talk more about a certain criterion in the textual reviews, it will be assigned a higher weight. We can measure it based on the cardinality of $t_{ij}^{k}$, which describes how many sentences the users employ to describe a criterion. Consequently, the weight for the criterion $C_j$ ($j = 1, \ldots, n$) is computed as follows:

$$\omega_j = \frac{\sum_{i=1}^{m} \sum_{k=1}^{K_i} N_{ij}^{k}}{\sum_{r=1}^{n} \sum_{i=1}^{m} \sum_{k=1}^{K_i} N_{ir}^{k}}, \; j = 1, \ldots, n. \tag{6}$$

### 4.5. Consensual collective opinion

At this stage, we can integrate the decision information from the previous steps to obtain a consensual STandR-BUI value representing all users' opinions.

To achieve this, we first combine the numerical rating and the sentiment score into a STandR-BUI preference. Then, a consensus process is performed to derive the agreed group opinions for each alternative across different criteria (see Algorithm 1). Specifically, we start by using Eq. (2) to calculate the STandR-BUI preferences $\langle x_{ij}^{k}; c_{ij}^{k}\rangle$ ($i = 1, \ldots, m, j = 1, \ldots, n, k = 1, \ldots, K_i$) from the numerical rating $r_{ij}^{k}$ ($i = 1, \ldots, m, j = 1, \ldots, n, k = 1, \ldots, K_i$) and the sentiment score $s_{ij}^{k}$ ($i = 1, \ldots, m, j = 1, \ldots, n, k = 1, \ldots, K_i$) for all users' ratings and textual reviews regarding the alternative $H_i$ ($i = 1, \ldots, m$) against the criteria $C_1, \ldots, C_n$. Afterward, for each alternative $H_i$, we employ the optimization model (3) on the users' STandR-BUI preferences $\langle x_{ij}^{k}; c_{ij}^{k}\rangle$ ($i = 1, \ldots, m; j = 1, \ldots, n, k = 1, 2, ..K_i$) to derive the modified individual preferences ${x'}_{ij}^{k}$ and the group collective value $x'_{ij}$. Here we consider that all the users' preferences have been given equal weight in the CRP, although other weights could be assigned according, for instance, to the helpfulness of the review or the users' profiles. Then, we utilize Eq. (4) to compute the individual certainty degrees ${c'}_{ij}^{k}$, and Eq. (5) provides the collective certainty degree $c'_{ij}$. Subsequently, we obtain the modified BUI values $\langle {x'}_{ij}^{k}; {c'}_{ij}^{k}\rangle, (i = 1, \ldots, m; j = 1, \ldots, n, k = 1, 2, ..K_i)$ and the collective preferences $\langle x'_{ij}; c'_{ij}\rangle, (i = 1, \ldots, m; j = 1, \ldots, n)$ of the alternatives against the criteria, which summarized in the STandR-BUI decision matrix, described in Table 1.

[3] https://textblob.readthedocs.io/.

**Algorithm 1** CRP for STandR-BUI

**Input:** Original ratings $r^k_{ij}$ and sentiment scores of the text reviews $s^k_{ij}$, consensus parameters $\varepsilon, \mu_0 \in [0,1]$, $A = (A_1, A_2) : J^n \to J$ BUI aggregation function such that $A_1$ is the weighted average operator, cost array $\xi^k_{ij}$, users' weights $w_k$, criteria weights $\omega_j$.

**Output:** users' modified opinions $< x'^k_{ij}; c'^k_{ij} >$ and collective opinion $< x'_{ij}; c'_{ij} >$ in STandR-BUI form.

1: Obtain STandR-BUI opinions by

$$< x^k_{ij}; c^k_{ij} > := \begin{cases} \phi^{-1}([\min\left\{r^k_{ij}, s^k_{ij}\right\}, \max\left\{r^k_{ij}, s^k_{ij}\right\}]) & \text{if } |r^k_{ij} - s^k_{ij}| < 1 \\ < \frac{1}{2}; 0 > & \text{otherwise} \end{cases}.$$

2: Apply the following MCC model to obtain users' agreed preferences $x'^k_{ij}$ and a collective preference $x'_{ij} = A_1(x'^1_{ij}, ..., x'^{K_i}_{ij})$:

$$\min_{x'^k_{ij} \in [0,1]} \sum_{k=1}^{K_i} \sum_{j=1}^{n} |x'^k_{ij} - x^k_{ij}|$$

$$s.t. \begin{cases} x'_{ij} = \frac{1}{K_i} \sum_{k=1}^{K_i} x'^k_{ij}, j = 1, ..., n \\ |x'^k_{ij} - x'_{ij}| \le \varepsilon, k = 1, \dots, K_i, j = 1, ..., n \\ 1 - \frac{1}{K_i} \sum_{k=1}^{K_i} \sum_{j=1}^{n} \omega_j |x'^k_{ij} - x'_{ij}| \ge \mu_0 \end{cases}$$

3: Compute the modified certainty degrees as

$$c'^k_{ij} = \hat{c}^k_{ij} |x - x'| + c^k_{ij}(1 - |x - x'|),$$

where

$$\hat{c}^k_{ij} = \begin{cases} c^k_{ij} \cdot (1 - x^k_{ij}) & \text{if } x'^k_{ij} = 0 \\ c^k_{ij} \cdot x^k_{ij} & \text{if } x'^k_{ij} = 1 \\ c^k_{ij} \cdot \min\left\{ \frac{x^k_{ij}}{x'^k_{ij}}, \frac{1-x^k_{ij}}{1-x'^k_{ij}} \right\} & \text{if } x'^k_{ij} \in ]0,1[ \end{cases}.$$

4: Compute the collective uncertainty degree $c'_{ij} = A_2(< x'^1_{ij}; c'^1_{ij} >, ..., < x'^K_{ij}; c'^K_{ij} >)$

### 4.6. Ranking of alternatives

Finally, the STandR-BUI decision matrix derived in the previous step is used to obtain an overall evaluation for each alternative. The overall evaluations $P_i (i = 1, 2, \dots, m)$ of the alternatives $H_i (i = 1, 2, \dots, m)$ are computed by using the following weighted BUIs aggregation:

$$\begin{aligned} P_i &= \langle x'_i; c'_i \rangle \\ &= A(\langle x'_{i1}; c'_{i1} \rangle, \dots, \langle x'_{in}; c'_{in} \rangle) \\ &= \langle M_\omega(x'_{i1}, \dots, x'_{in}), 1 - \mathcal{L}([M_\omega(a_{i1}, \dots, a_{in}), M_\omega(b_{i1}, \dots, b_{in})]) \rangle, \end{aligned}$$

where $M_\omega$ is the weighted average with the weight of the criteria $\omega$ and $[a_{ij}, b_{ij}] = \phi(\langle x'_{ij}; c'_{ij} \rangle), (i = 1, \dots, m, j = 1, \dots, n)$. Finally, we rank the alternatives $P_i$ $(i = 1, \dots, m)$ based on their overall preference values $x'_i$ $(i = 1, \dots, m)$. The ranking is determined by the principle that a higher overall preference value corresponds to a better rank, with alternatives being ordered accordingly.

## 5. Case study

In this section, we illustrate the proposed consensual framework for fusing rating and textual information via a practical example of the evaluation of hotels.

### 5.1. Problem and datasets

As our framework is developed to combine the numerical ratings and textual reviews, we attempt to find a dataset for hotel reviews in which both numerical ratings and textual reviews against different

**Table 1**
Collective STandR-BUI decision matrix obtained after CRP.

| Alternatives | $C_1$ | $C_2$ | ⋯ | $C_n$ |
|---|---|---|---|---|
| $H_1$ | $\langle x'_{11}; c'_{11} \rangle$ | $\langle x'_{12}; c'_{12} \rangle$ | ⋯ | $\langle x'_{1n}; c'_{1n} \rangle$ |
| $H_2$ | $\langle x'_{21}; c'_{21} \rangle$ | $\langle x'_{22}; c'_{22} \rangle$ | ⋯ | $\langle x'_{2n}; c'_{2n} \rangle$ |
| ⋮ | ⋮ | ⋮ | ⋮ | |
| $H_m$ | $\langle x'_{m1}; c'_{m1} \rangle$ | $\langle x'_{m2}; c'_{m2} \rangle$ | ⋯ | $\langle x'_{mn}; c'_{mn} \rangle$ |

assessment criteria are present. For this purpose, we utilize the Tripadvisor datasets[4] of four Turkish hotels $H_i, (i = 1, 2, 3, 4)$ that contains similar evaluation information.

In the review process, the users are asked to rate the hotel regarding five predefined criteria, namely, *value ($C_1$), room ($C_2$), service ($C_3$), cleanliness($C_4$), and location($C_5$)* using a five points Likert scale (1–5). The users are also asked to share the textual account of their experience regarding the criteria. We denote by $r^k_{ij}$ the rating given by the $k$th user for criterion $C_j$. The user might not mention certain criteria in their textual review or provide a numerical rating, and this is evident in the dataset.

### 5.2. Application of the proposed framework

We now employ the proposed decision-making framework for evaluating hotels to obtain ranking and operational insights. First, the criteria detection tool is applied to the textual review ($T^k_i$, $i = 1, 2, 3, 4, k = 1, \dots, K_i$) of each user for detecting the texts that he/she used to describe the experience about different criteria. Subsequently, we obtain the sentences that describe the user's view of each criterion $t^k_{ij} = \{\tau^k_{ij1}, \dots, \tau^k_{ijN^k_{ij}}\}, (j = 1, \dots, 5)$. For example, the review given by the $k$th user $U^i_k$ for the hotel $H_i$ was $T^k_i = \{$*The hotel is in a great location with only a 20 min transfer from Antalya airport. The view was very beautiful. In turkyie was very hot but sometimes here is nice windy. Food is very good. The staff was very friendly. Overall the hotel is very clean and tidy. Our room was good with comfortable beds. It is a very good hotel. We recommend this hotel to everybody.*$\}$. This review was segregated according to the criteria detection tool as follows:

- $t^k_{i1}$ (value): It is a very good hotel. We recommend this hotel to everybody.
- $t^k_{i2}$ (room): Our room was good with comfortable beds.
- $t^k_{i3}$ (services): Food is very good. The staff was very friendly.
- $t^k_{i4}$ (cleanliness): Overall the hotel is very clean and tidy
- $t^k_{i5}$ (location): The hotel is in a great location with only a 20 min transfer from Antalya airport. The view was very beautiful. In turkyie was very hot but sometimes here is nice and windy.

As many users do not provide ratings and textual reviews against all the criteria, we refined the dataset considering only the users who provide ratings and textual reviews concerning the five criteria. After filtering the dataset with this criterion, the hotels $H_i (i = 1, 2, 3, 4)$ have the following number of evaluations by users: $K_1 = 25$, $K_2 = 86$, $K_3 = 50$, $K_4 = 42$. This refined dataset was subsequently used in the hotels' evaluation problem, which can be considered as an LSGDM problem.

Next, we analyze the sentiment for each criterion $C_j$ based on the textual descriptions $t^k_{ij} = \{\tau^k_{ij1}, \dots, \tau^k_{ijN^k_{ij}}\}$ provided by the $k$th user for hotel $H_i$. For each $t^k_{ij}$, we apply the procedure described in the previous section to calculate the sentiment probability values $(p_-, p_., p_+)^k_{ij} = \left( \frac{N^k_{ij-}}{N^k_{ij}}, \frac{N^k_{ij.}}{N^k_{ij}}, \frac{N^k_{ij+}}{N^k_{ij}} \right)$, where $N^k_{ij-}$, $N^k_{ij.}$, and $N^k_{ij+}$ represent the counts of negative, neutral, and positive sentiments, respectively, for criterion

[4] https://www.kaggle.com/datasets/muhammadhassanniazi/tripadvisor-hotels-dataset.

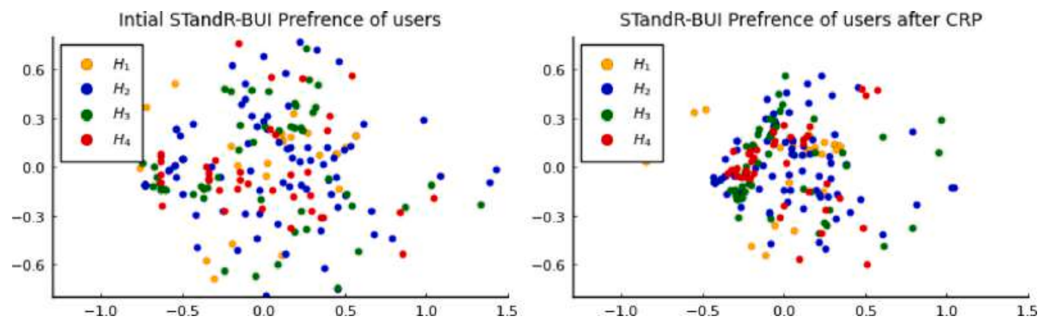



**Fig. 8.** Multi-dimensional scaling representation of the distances between users' initial and modified opinions.

**Table 2**
The computation of sentiment probability from the text review.

| Texts ($t^k_{i3}$) | Sentiment ($Senti(\tau^k_{i3l})$) | Probability $(p_-, p_., p_+)^k_{i3}$ |
|---|---|---|
| Buffet breakfast is fantastic. | $s_+$ | |
| Dinner experience at the restaurant was superb and really a five star. | $s_+$ | |
| However, I was not impressed by my dishes though at the 14th floor bar. | $s_-$ | $(\frac{N^k_{i3-}}{N^k_{i3}}, \frac{N^k_{i3.}}{N^k_{i3}}, \frac{N^k_{i3+}}{N^k_{i3}}) = (\frac{1}{4}, \frac{1}{4}, \frac{2}{4})$ |
| Room service experience was also just so so. | $s_.$ | |

**Table 3**
Hotels' collective opinions in STandR-BUI after CRP.

| Hotels | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|---|
| $H_1$ | $\langle 0.659; 0.763\rangle$ | $\langle 0.894; 0.847\rangle$ | $\langle 0.691; 0.781\rangle$ | $\langle 0.952; 0.873\rangle$ | $\langle 0.647; 0.778\rangle$ |
| $H_2$ | $\langle 0.595; 0.75\rangle$ | $\langle 0.691; 0.771\rangle$ | $\langle 0.703; 0.779\rangle$ | $\langle 0.931; 0.906\rangle$ | $\langle 0.637; 0.729\rangle$ |
| $H_3$ | $\langle 0.554; 0.752\rangle$ | $\langle 0.587; 0.714\rangle$ | $\langle 0.566; 0.756\rangle$ | $\langle 0.906; 0.84\rangle$ | $\langle 0.559; 0.653\rangle$ |
| $H_4$ | $\langle 0.516; 0.7\rangle$ | $\langle 0.549; 0.69\rangle$ | $\langle 0.605; 0.741\rangle$ | $\langle 0.938; 0.899\rangle$ | $\langle 0.647; 0.728\rangle$ |

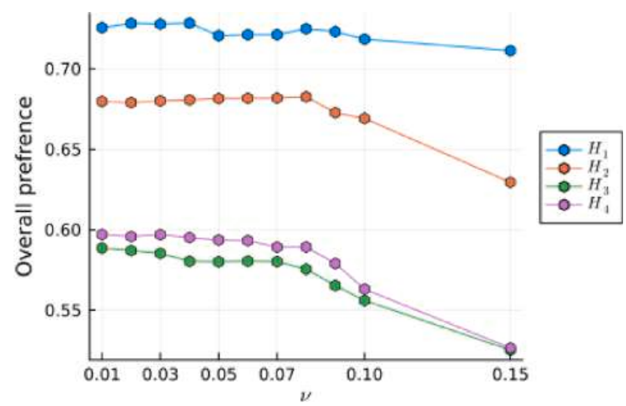



**Fig. 9.** Representation of the performance variations of each alternative with respect to the support parameter $\nu$.

$C_j$ as assessed by user $U_k$. For example, Table 2 illustrates how these probability values are computed for $t^k_{i3}$, the text review related to the service criterion $C_3$ for hotel $H_i$ provided by user $U_k$. This process is repeated for each textual review across all criteria and users to obtain the sentiment probabilities.

In this way, the numerical rating against each criterion and the sentiment score of each review for all the criteria are obtained. Now, we proceed to encapsulate the rating and sentiment preference information into a single value via STandR-BUI utilizing the Eqs. (1) and (2). For example, consider the sentiment evaluation of service criteria described in Table 2 with sentiment probability $(p_-, p_., p_+)^k_{ij} = (\frac{1}{4}, \frac{1}{4}, \frac{2}{4})$ and normalized rating $r^k_{ij} = \frac{4}{5}$ on the $[0, 1]$ scale. Utilizing Eq. (1), we compute the sentiment score $s^k_{ij} = Sc((\frac{1}{4}, \frac{1}{4}, \frac{2}{4})) = 0.595$. Subsequently, we obtain the STandR-BUI representation with the help of Eq. (2) as $\langle x^k_{ij}; c^k_{ij}\rangle = \langle 0.111, 0.565\rangle$. Similarly, we transform the sentiment and rating into STandR-BUI for all users' preferences.

Next, we assign weights to the criteria based on the concerns expressed by users in their text reviews. Using Eq. (6), we compute the weights for the criteria as $\omega = (0.287, 0.145, 0.407, 0.076, 0.085)$.

To obtain the STandR-BUI group preferences for each hotel $H_i (i = 1, 2, 3, 4)$ and criteria $C_j (j = 1, 2, 3, 4, 5)$ from their STandR-BUI preference information, the CRP described in Algorithm 1 is applied. As a result, we derive the consensual STandR-BUI preferences of users for different hotels, which are depicted in Fig. 8 via multidimensional scaling (Cox & Cox, 2008). Further, the group consensus opinions for each alternative concerning the criteria are shown in Table 3. Note that all the users' have been given equal weight in the CRP.

From Table 3, we compute the overall preference $P_i = \langle x'_i; c'_i\rangle, (i = 1, 2, 3, 4)$ of each hotel $H_i (i = 1, 2, 3, 4)$ in terms of STandR-BUI values by aggregating with respect to the criteria using 4.6. The following results were obtained: $P_1 = \langle 0.7274; 0.7921\rangle$, $P_2 = \langle 0.6820; 0.7748\rangle$, $P_3 = \langle 0.5907; 0.7461\rangle$, $P_4 = \langle 0.5915; 0.7292\rangle$. Finally, based on the overall degree of preference of the hotels represented by STandR-BUIs, we obtain the following agreed ranking of hotels: $H_1 \succ H_2 \succ H_4 \succ H_3$.

### 5.3. Sensitivity analysis

In this section, we investigate the robustness of the final decision outcome obtained by applying our proposed decision-making methodology in the previous section considering the variation in the key decision-making parameters.

#### 5.3.1. Shape parameter of sentiment labels

Here, we analyze the impact of the shape parameter of the sentiment labels ($\nu$) on the overall performance of the alternatives. The parameter $\nu$ essentially controls the support of the sentiment label "neutrality". Further increasing the value of $\nu$ will shrink the support for the positive and negative sentiments. To understand its impact, we compute the overall performance of the alternatives for different values of $\nu$, keeping other parameter values and dataset unchanged as per the original experiment in Section 5. The changes in the alternatives' overall performance are reported in Fig. 9.

We observe that the overall performances of the alternatives slightly change with small variations in support, while the rankings of the alternatives remain the same in all cases. Additionally, the changes in $H_1$'s overall performance remain small even for a high value of $\nu = 0.15$. This possibly occurs due to the users' opinions regarding $H_1$ being extremely positive on the bipolar scale, which restricts the opinions from shifting from positive sentiment labels to neutral sentiment. On the other hand, some opinions related to $H_2$, $H_3$, and $H_4$ are moderately positive. Therefore, the increase in values of $\nu$ enables them to switch to neutral sentiment labels and subsequently lower the score. For this reason, the overall performances of $H_2$, $H_3$, and $H_4$ decrease significantly for high values of $\nu$.

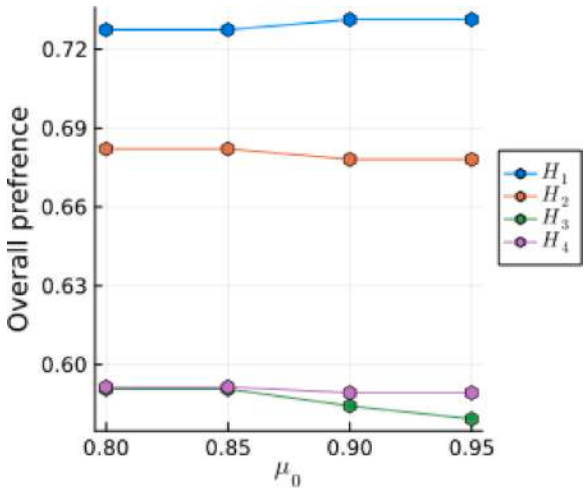



**Fig. 10.** Representation of the performance variations of each alternative with respect to the consensus parameter $\mu_0$.

The above analysis suggests that the support parameter of sentiment labels could be crucial to the robustness of the results when the sentiments of opinions about the alternatives are distributed around a neutral point. In such cases, it is necessary to analyze the impact of the shape parameter on the final decisions to ensure robustness. When the sentiments in the text reviews are extremely positive or negative, small variations in the shape parameter do not significantly impact the final results.

#### *5.3.2. Consensus threshold*

Here, we investigate the impact of the consensus threshold ($\mu_0$) on the overall performance of the alternatives. For this purpose, we kept the maximum allowable distance between the users' opinion and the group opinion as $\varepsilon = 0.2$ and computed the overall performance of the alternatives for different values of $\mu_0$ for the aforementioned hotels' evaluation problem with the same initial parameters. The changes in the overall performances of the alternatives for the consensus threshold are depicted in Fig. 10. We observe that the overall performance values do not change for $\mu_0 \in [0.8, 0.85]$. This behavior could be explained through the relationship of the parameters $\varepsilon$ and $\mu_0$ and their impact on the optimal solution of the optimization model (3) in line with the study by García-Zamora, Dutta, Massanet et al. (2023). As the vertices of the polytope formed by the $\varepsilon = 0.2$ constraints of the optimization model (3) belong to the boundary of the regions formed by the $\mu_0$ constraint of such model, the consensus opinions provided by the optimization model did not change, and subsequently, the overall performance of the alternatives for consensus thresholds $\mu_0 = 0.8$ and $\mu_0 = 0.85$ remains unaltered. When the consensus threshold takes the values $\mu_0 = 0.9$ and $\mu_0 = 0.95$, the users' modified opinions obtained from model (3) change for all the alternatives, but group opinions remain unaltered for the alternatives $H_1, H_2$, and $H_4$. As a consequence, these alternatives' overall performances remain unaltered for these consensus threshold values.

The above analysis demonstrates that the consensus threshold ($\mu_0$) and the maximum allowable distance ($\varepsilon$) interact and impact the final outcome only under certain specific conditions. However, the choice of these values should be guided by the context of the decision-making process. A very high value of $\mu_0$ could force significant modification of the original opinions, potentially misrepresenting true preferences. Conversely, a low consensus threshold could help preserve original opinions but may not provide a sufficiently agreed-upon view. Therefore, the chosen values should strike a balance between these aspects to achieve broad acceptance.

**Table 4**
Computational time cost (in seconds) to obtain the final ranking of the alternatives using the proposed LSGDM framework for different numbers of users.

| Number of users | Average time cost (s) | Standard deviation |
|---|---|---|
| 50 | 3.2238 | 0.0454 |
| 100 | 6.3018 | 0.0129 |
| 200 | 12.6224 | 0.0153 |
| 500 | 31.5736 | 0.0212 |
| 1000 | 63.3469 | 0.0422 |
| 2000 | 127.3636 | 0.0850 |

## 6. Evaluation of the model

This section attempts to show the technical advantages of our proposal. As our proposal focuses on an LSGDM framework for managing text reviews and rating information, we emphasize the general approaches that LSGDM follows to showcase the advantages of the new decision-making framework. In this context, we consider three key aspects to evaluate the proposed framework: practical feasibility, modeling impact, and comparison with existing methods. The practical feasibility aspect assesses the framework's capability to handle real-life LSGDM scenarios with more than 50 decision-makers (DMs), ensuring efficient operation within reasonable timeframes. This aspect demonstrated through computational time complexities in Section 6.1, is essential for the method's real-world applications. The modeling impact aspect focuses on how the proposed model influences final decision outcomes, highlighting its effectiveness in tackling complex decision problems. This aspect is discussed in Section 6.2.1. Finally, comparison with existing methods allows us to evaluate the proposed framework's performance relative to similar methods and to examine how different information fusion mechanisms and modeling approaches impact decision results, as detailed in Section 6.2.2.

### *6.1. Computational time analysis*

This subsection demonstrates the computational aspects of applying our proposed decision-making framework in real life by simulating its performance in several scenarios.

Let us describe the experiment setup to evaluate the computational time. Although our framework allows the number of users could be different for the evaluation of the different alternatives, we assume that each alternative is evaluated by the same number of users. Further, we set the length of the text review to 625 words, which was the average length in our dataset. With this view in mind, we generate a random problem instance with four alternatives and five criteria (aspects/features). Each alternative will be evaluated by $K$ users via a textual review and a numerical rating against five predefined criteria. The consensus threshold is set as $\mu_0 = 0.8$, while the maximum allowable distance between the users and the group is $\varepsilon = 0.2$. Given this context, our decision-making framework is executed using Python 3.9.12 and Julia 1.7.3 on a Windows 10 Professional desktop. The system is equipped with a 2.5 GHz Intel Core i7-11700 CPU and 16 GB of RAM. We conduct experiments with varying numbers of users and present the average time taken by our algorithm to finalize alternative rankings over 10 repetitions in Table 4 along with the standard deviation. The evidence suggests a gradual increase in mean time cost with an increasing number of users. Furthermore, it is noteworthy that the proposed decision-making framework maintains its effectiveness even when dealing with larger-scale challenges, as exemplified by its suitability for scenarios involving 2000 users. Additionally, it demonstrates the capability to derive the final decision outcome in a reasonable time while maintaining a minimal standard deviation in time costs. Moreover, most of the time, the algorithm is spent processing text reviews, particularly in aspect detection and converting text reviews and ratings to BUI values.

**Table 5**
Overall scores of the alternatives considering the rating and text reviews separately.

| $a$ | Overall rating score | Overall sentiment score |
|---|---|---|
| $a_1$ | 0.9317 | 0.8943 |
| $a_2$ | 0.9050 | 0.8497 |
| $a_3$ | 0.9100 | 0.7538 |
| $a_4$ | 0.9467 | 0.7430 |

**Table 6**
Final ranking of the alternatives.

| Method | Final result |
|---|---|
| Liao et al. (2023) | $H_1 \succ H_3 \succ H_4 \succ H_2$ |
| Proposed method | $H_1 \succ H_2 \succ H_4 \succ H_3$ |

### 6.2. Comparison analysis

Below, we compare our proposal with other approaches. First, we analyze the impact of STandR-BUI information when modeling the users' preferences, in comparison with using either only ratings or only text reviews. Then, we compare our framework with the recently published work of Liao et al. (2023), who provided a decision model that also integrates text and rating reviews.

#### 6.2.1. On BUI modeling perspective

Classical data-driven proposals that apply LSGDM techniques are usually constrained to either only use ratings (Garcí a-Zamora, Dutta, Labella et al., 2023) or the sentiment of the text (Zuheros et al., 2021, 2023). To observe the effect of considering the rating and sentiment together over the ranking of the alternatives, we attempt to compute the ranking orders of the alternatives by considering the numerical rating and text reviews separately.

To compute the ranking from the ratings provided by reviewers against the five mentioned criteria, we first employ the MCC model over the normalized ratings and obtain the group consensus opinions for each hotel. Afterward, we compute the alternatives' overall scores using the weighted average. The overall scores of the alternatives are depicted in Table 5. Similarly, we consider only the textual reviews given by the reviewers to compute the ranking of the alternatives. As described earlier, from the reviewer's text review against a criterion, we obtain the sentiment probability and subsequently the sentiment score. After that, we employ the MCC model to find the consensual group sentiment score of the alternatives against the criteria. Finally, by aggregating the group sentiment scores across the criteria with the weighted arithmetic mean, we obtain the overall sentiment score of the alternatives, which are reported in Table 5

Based on the overall rating score, we obtain the following ranking order of the alternatives: $H_4 \succ H_1 \succ H_3 \succ H_2$. On the other hand, the overall sentiment score produces the following rank order of the alternatives: $H_1 \succ H_2 \succ H_3 \succ H_4$. Note that the numerical rating finds $H_4$ as the best hotel, while the sentiment provides $H_1$ as the best alternative. Further, the numerical rating and sentiment score produced ranking orders different from BUI-based ranking, which combined rating and sentiment. Therefore, considering both aspects (numerical and textual review) during the evaluation considerably impacts the results.

#### 6.2.2. Comparison with an existing study

In this section, we develop a fair comparison of our data-driven decision-making framework with a recently published study that offers a different mechanism for processing text and rating reviews. Concretely, our proposal is compared with the unifying text and rating mechanism by Liao et al. (2023) for its striking similarity in terms of input information and information unification mechanism with our proposal. For that purpose, we implement the algorithm provided by Liao et al. (2023) in their study and apply it to our dataset. As Liao et al. (2023) does not consider the weight of the attributes, we kept the equal weight in our settings to remove the influence of the attribute's weight. Table 6 reported the final ranking order of the alternatives. Note that both approaches identified $H_1$ as the best alternative. However, the ranking of the rest of the alternatives differs.

Two potential reasons contribute to the difference in ranking. Firstly, the process of unifying text and ratings via a shared utility scale assumes that the expected value of ratings across specific aspects aligns with the corresponding expected utility derived from the textual content. In other words, Liao et al. (2023) assume complete consistency between rating and text, which is not quite rational when considering online review information that is inherently inconsistent. On the other hand, our framework acknowledges and accommodates this inconsistency between text and ratings through the incorporation of uncertainty within the STandR-BUI unification process.

Secondly, the stochastic dominance method used to rank the alternatives by Liao et al. (2023) could neglect information under certain criteria, for not establishing a dominance relation, in the overall dominance computation between the alternatives, and consequently affecting the final ranking of the alternatives. In contrast, our straightforward aggregation-based information fusion mechanism appropriately considers all attribute information, thereby ensuring a more comprehensive evaluation.

Additionally, the decision-making framework proposed by Liao et al. (2023) involves numerous parameters, including the weight of text and rating information, three behavioral parameters to define the value function, the number of evaluation terms, and weights associated with different dominance relations. This abundance of parameters complicates the task of selecting appropriate values during application, adding to the overall complexity. In contrast, our algorithm operates with simplicity, relying solely on consensus model parameters and shape-controlling parameters of the sentiment labels.

## 7. Conclusions

Microblogging services provide a valuable source of information for decision-making processes, but existing approaches often overlook numerical ratings, leading to an oversimplified representation of users' opinions. This paper has introduced the STandR-BUI preference-modeling structure, efficiently summarizing sentiment scores and numerical ratings into a single BUI value. Additionally, a data-driven framework based on LSGDM is proposed to manage users' reviews given as ratings and text, incorporating an automatic consensus model for processing thousands of reviews to obtain a collective opinion on performance across multiple criteria. Such a consensus model addresses the multifaceted nature of users' reviews, weighting diverse viewpoints to foster more inclusive and representative decisions.

Even though our approach to detecting aspects in textual reviews is applicable in a wide number of practical cases, it has some limitations. Regarding Stanford CoreNLP, when we are dealing with a large number of textual reviews, it may require substantial resources, including memory and processing power. Further, the accuracy of Stanford CoreNLP's dependency parser may not be as high as some state-of-the-art parsers, which can affect the performance of downstream NLP tasks that rely on accurate parsing. Furthermore, we are cognizant of the limited language support of both Stanford CoreNLP and TextBlob. To address this, we have taken measures to carefully preprocess and filter the data to focus on the relevant languages in our dataset (only English in this study). However, these limitations could be overcome by using NLP libraries that offer more straightforward APIs and pre-trained models for specific tasks (aspects detection and SA) such as SpikeGPT (Zhu et al., 2023) or LLaMA (Touvron et al., 2023). Other possible improvements may focus on refining the consensus model to handle various types of uncertainties and investigating the integration of additional data sources to further improve the accuracy. Moreover, exploring the generalization of this approach to other decision-making methodologies and examining its performance in real-world applications will be valuable avenues for research.

### CRediT authorship contribution statement

**Diego García-Zamora:** Conceptualization, Formal analysis, Methodology, Visualization, Software, Writing – original draft, Writing – review & editing. **Bapi Dutta:** Conceptualization, Formal analysis, Methodology, Visualization, Software, Writing – original draft, Writing – review & editing. **LeSheng Jin:** Validation, Writing – review & editing. **Zhen-Song Chen:** Validation, Writing – review & editing. **Luis Martínez:** Conceptualization, Validation, Funding acquisition, Project administration, Writing – review & editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This work is partially supported by the Grants for the Requalification of the Spanish University System for 2021–2023 in the María Zambrano modality (UJA13MZ).

### Data availability

Used data link has been provided in the manuscript.

### References

Agarwal, A. K., Mahoney, K., Lanza, A. L., Klinger, E. V., Asch, D. A., Fausti, N., Tufts, C., Ungar, L., & Merchant, R. M. (2019). Online ratings of the patient experience: emergency departments versus urgent care centers. *Annals of Emergency Medicine*, *73*(6), 631–638.

Agarwal, A. K., Wong, V., Pelullo, A. M., Guntuku, S., Polsky, D., Asch, D. A., Muruako, J., & Merchant, R. M. (2020). Online reviews of specialized drug treatment facilities—identifying potential drivers of high and low patient satisfaction. *Journal of General Internal Medicine*, *35*(6), 1647–1653.

Ben-Arieh, D., & Easton, T. (2007). Multi-criteria group consensus under linear cost opinion elasticity. *Decision Support Systems*, *43*(3), 713–721.

Borg, A., & Boldt, M. (2020). Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Systems with Applications*, *162*, Article 113746. http://dx.doi.org/10.1016/j.eswa.2020.113746.

Butler, C., & Rothstein, A. (1987). *On conflict and consensus: A handbook on formal consensus decision making*. Portland: Food Not Bombs Publishing.

Cao, J., Xu, X., Yin, X., & Pan, B. (2022). A risky large group emergency decision-making method based on topic sentiment analysis. *Expert Systems with Applications*, *195*, Article 116527. http://dx.doi.org/10.1016/j.eswa.2022.116527.

Carvalho, G., Vivacqua, A. S., Souza, J. M., & Medeiros, S. P. J. (2011). Lasca: A large scale group decision support system. *Journal of Universal Computer Science*, *17*(2), 261–275.

Chang, Y.-C., Ku, C.-H., & Chen, C.-H. (2019). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management*, *48*, 263–279.

Cho, H. S., Sosa, M. E., & Hasija, S. (2022). Reading between the stars: Understanding the effects of online customer reviews on product demand. *Manufacturing & Service Operations Management*, *24*(4), 1977–1996.

Cox, M. A. A., & Cox, T. F. (2008). Multidimensional scaling. In *Handbook of data visualization* (pp. 315–347). Berlin, Heidelberg: Springer Berlin Heidelberg.

Devgan, L. L., Klein, E. J., Fox, S., & Ozturk, T. (2020). Bifurcation of patient reviews: An analysis of trends in online ratings. *Plastic and Reconstructive Surgery-Global Open*, *8*(4), Article e2781. http://dx.doi.org/10.1097/GOX.0000000000002781.

Dong, J., Chen, Y., Gu, A., Chen, J., Li, L., Chen, Q., Li, S., & Xun, Q. (2020). Potential trend for online shopping data based on the linear regression and sentiment analysis. *Mathematical Problems in Engineering*, *2020*(1), Article 4591260. http://dx.doi.org/10.1155/2020/4591260.

Du, J., Zheng, L., He, J., Rong, J., Wang, H., & Zhang, Y. (2020). An interactive network for end-to-end review helpfulness modeling. *Data Science and Engineering*, *5*(3), 261–279.

García-Zamora, D., Dutta, B., Labella, Á., & Martínez, L. (2023). A fuzzy-set based formulation for minimum cost consensus models. *Computers & Industrial Engineering*, *181*, Article 109295. http://dx.doi.org/10.1016/j.cie.2023.109295.

García-Zamora, D., Dutta, B., Massanet, S., Riera, J. V., & Martínez, L. (2023). Relationship between the distance consensus and the consensus degree in comprehensive minimum cost consensus models: A polytope-based analysis. *European Journal of Operational Research*, *306*(2), 764–776.

García-Zamora, D., Labella, Á., Ding, W., Rodríguez, R. M., & Martínez, L. (2022). Large-scale group decision making: A systematic review and a critical analysis. *IEEE/CAA Journal of Automation Sinica*, *9*(6), 949–966.

He, S., & Wang, Y. (2023). Evaluating new energy vehicles by picture fuzzy sets based on sentiment analysis from online reviews. *Artificial Intelligence Review*, *56*(3), 2171–2192.

Heidary Dahooie, J., Raafat, R., Qorbani, A. R., & Daim, T. (2021). An intuitionistic fuzzy data-driven product ranking model using sentiment analysis and multi-criteria decision-making. *Technological Forecasting and Social Change*, *173*, Article 121158. http://dx.doi.org/10.1016/j.techfore.2021.121158.

Hermida, A. (2010). Twittering the news: The emergence of ambient journalism. *Journalism Practice*, *4*(3), 297–308.

Herrera-Viedma, E., Cabrerizo, F. J., Kacprzyk, J., & Pedrycz, W. (2014). A review of soft consensus models in a fuzzy environment. *Information Fusion*, *17*, 4–13.

Herrera-Viedma, E., Herrera, F., & Chiclana, F. (2002). A consensus model for multiperson decision making with different preference structures. *IEEE Transactions on Systems, Man & Cybernetics, Part A (Systems & Humans)*, *32*, 394–402.

Hlee, S. (2020). How reviewer level affects review helpfulness and reviewing behavior across hotel classifications: The case of Seoul in Korea. *Industrial Management & Data Systems*, *121*(6), 1191–1215.

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *vol. 8*, In *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 216–225). Association for the Advancement of Artificial Intelligence (AAAI).

Jerripothula, K. R., Rai, A., Garg, K., & Rautela, Y. S. (2020). Feature-level rating system using customer reviews and review votes. *IEEE Transactions on Computational Social Systems*, *7*(5), 1210–1219.

Jia, S. (2020). Motivation and satisfaction of Chinese and US tourists in restaurants: A cross-cultural text mining of online reviews. *Tourism Management*, *78*, Article 104071. http://dx.doi.org/10.1016/j.tourman.2019.104071.

Jin, L., Mesiar, R., Borkotokey, S., & Kalina, M. (2018). Certainty aggregation and the certainty fuzzy measures. *International Journal of Intelligent Systems*, *33*(4), 759–770.

Kacprzyk, J., & Fedrizzi, M. (1988). A 'soft' measure of consensus in the setting of partial (fuzzy) preferences. *European Journal of Operational Research*, *34*(3), 316–325.

Karthik, R., & Ganapathy, S. (2021). A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce. *Applied Soft Computing*, *108*, Article 107396. http://dx.doi.org/10.1016/j.asoc.2021.107396.

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

Labella, Á., Liu, H., Rodríguez, R. M., & Martinez, L. (2020). A cost consensus metric for consensus reaching processes based on a comprehensive minimum cost model. *European Journal of Operational Research*, *281*(2), 316–331.

Liang, D., Dai, Z., Wang, M., & Li, J. (2020). Web celebrity shop assessment and improvement based on online review with probabilistic linguistic term sets by using sentiment analysis and fuzzy cognitive map. *Fuzzy Optimization and Decision Making*, *19*, 561–586.

Liao, H., Wang, J., & Xu, Z. (2023). Unifying star ratings and text reviews in linguistic terms for product competitiveness analysis based on stochastic dominance. *IEEE Transactions on Computational Social Systems*, *11*(3), 3678–3690.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014). The stanford coreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55–60).

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113.

Mesiar, R., Borkotokey, S., Jin, L., & Kalina, M. (2018). Aggregation under uncertainty. *IEEE Transactions on Fuzzy Systems*, *26*(4), 2475–2478.

Palomares, I., Estrella, F. J., Martínez, L., & Herrera, F. (2014). Consensus under a fuzzy context: Taxonomy, analysis framework afryca and experimental case of study. *Information Fusion*, *20*, 252–271.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *proceedings of the 2002 conference on empirical methods in natural language processing (EMNLP 2002)* (pp. 79–86). Association for Computational Linguistics.

Qin, J., Zeng, M., Wei, X., & Pedrycz, W. (2023). Ranking products through online reviews: A novel data-driven method based on interval type-2 fuzzy sets and sentiment analysis. *Journal of the Operational Research Society*, *75*(5), 860–873.

Rodríguez, R. M., Labella, Á., Dutta, B., & Martínez, L. (2021). Comprehensive minimum cost models for large scale group decision making with consistent fuzzy preference relations. *Knowledge-Based Systems*, *215*, Article 106780. http://dx.doi.org/10.1016/j.knosys.2021.106780.

Sharma, S. S., & Dutta, G. (2021). SentiDraw: Using star ratings of reviews to develop domain specific sentiment lexicon for polarity determination. *Information Processing & Management*, *58*(1), Article 102412. http://dx.doi.org/10.1016/j.ipm.2020.102412.

Shi, J. C., Yang, Y., & Qiu, S. (2021). Making sales strategies based on the existing shopping reviews. *Journal of Physics: Conference Series*, *1865*, Article 042047. http://dx.doi.org/10.1088/1742-6596/1865/4/042047.

Song, H.-H., Dutta, B., García-Zamora, D., Wang, Y.-M., & Martínez, L. (2024). Managing non-cooperative behaviors in consensus reaching process: A novel multi-stage linguistic LSGDM framework. *Expert Systems with Applications*, *240*, Article 122555. http://dx.doi.org/10.1016/j.eswa.2023.122555.

Sueur, C., Deneubourg, J. L., & Petit, O. (2012). From social network (centralized vs. decentralized) to collective decision-making (unshared vs. shared consensus). *PLoS One*, *7*(2), Article e32566. http://dx.doi.org/10.1371/journal.pone.0032566.

Tang, C., Mehl, M. R., Eastlick, M. A., He, W., & Card, N. A. (2016). A longitudinal exploration of the relations between electronic word-of-mouth indicators and firms' profitability: findings from the banking industry. *International Journal of Information Management*, *36*(6), 1124–1132.

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014). Learning sentiment-specific word embedding for Twitter sentiment classification. In K. Toutanova, & H. Wu (Eds.), *proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 1555–1565). Baltimore, Maryland: Association for Computational Linguistics.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. arXiv:2302.13971.

Wu, X., & Liao, H. (2021). Modeling personalized cognition of customers in online shopping. *Omega*, *104*, Article 102471. http://dx.doi.org/10.1016/j.omega.2021.102471.

Xu, Y. J., Wen, X. W., & Zhang, W. C. (2018). A two-stage consensus method for large-scale multi-attribute group decision making with an application to earthquake shelter selection. *Computers & Industrial Engineering*, *116*, 113–129.

You, L., Han, F., Peng, J., Jin, H., & Claramunt, C. (2022). ASK-RoBERTa: A pretraining model for aspect-based sentiment classification via sentiment knowledge mining. *Knowledge-Based Systems*, *253*, Article 109511. http://dx.doi.org/10.1016/j.knosys.2022.109511.

Zhang, Z., Guo, C., & Martínez, L. (2017). Managing multigranular linguistic distribution assessments in large-scale multiattribute group decision making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *47*(11), 3063–3076.

Zhang, D., Wu, C., & Liu, J. (2020). Ranking products with online reviews: A novel method based on hesitant fuzzy set and sentiment word framework. *Journal of the Operational Research Society*, *71*(3), 528–542.

Zhu, B., Guo, D., & Ren, L. (2022). Consumer preference analysis based on text comments and ratings: A multi-attribute decision-making perspective. *Information & Management*, *59*(3), Article 103626. http://dx.doi.org/10.1016/j.im.2022.103626.

Zhu, R.-J., Zhao, Q., Li, G., & Eshraghian, J. K. (2023). SpikeGPT: Generative pre-trained language model with spiking neural networks. arXiv:2302.13939.

Zuheros, C., Martinez-Camara, E., Herrera-Viedma, E., & Herrera, F. (2021). 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. *Information Fusion*, *68*, 22–36.

Zuheros, C., Martinez-Camara, E., Herrera-Viedma, E., & Herrera, F. (2023). Crowd decision making: Sparse representation guided by sentiment analysis for leveraging the wisdom of the crowd. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *53*, 369–379.