Information Fusion 97 (2023) 101821

Contents lists available at ScienceDirect

Information Fusion

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

Full length article

Explainable Crowd Decision Making methodology guided by expert natural language opinions based on Sentiment Analysis with Attention-based Deep Learning and Subgroup Discovery

Cristina Zuheros [a],[*], Eugenio Martínez-Cámara [a], Enrique Herrera-Viedma [a], Iyad A. Katib [b], Francisco Herrera [a]

[a] Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18071, Granada, Spain

[b] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Keywords:
Crowd decision making
Explainability
Attention mechanisms
Subgroup discovery
Social media opinions

ABSTRACT

There exist a high demand to provide explainability to artificial intelligence systems, where decision making models are included. This paper focuses on crowd decision making using natural language evaluations from social media with the aim to provide explainability. We present the Explainable Crowd Decision Making based on Subgroup Discovery and Attention Mechanisms (ECDM-SDAM) methodology as an a posteriori explainable process that captures the wisdom of crowds that is naturally provided in social media opinions. It extracts the opinions from social media texts using a deep learning based sentiment analysis approach called Attention based Sentiment Analysis Method. The methodology includes a backward process that provides explanations to justify its sense-making procedure by applying mainly the attention mechanism on texts and subgroup discovery on opinions. We evaluate the methodology in the real case study of the TripR-2020Large dataset for restaurant choice. The results show that the ECDM-SDAM methodology provides easy understandable explanations that elucidates the key reasons that support the output of the decision process.

1. Introduction

Decision Making (DM) models support people in making a decision. They establish a ranking of alternatives based on the evaluations of a group of experts, usually to get the best alternative [1]. *Crowd Decision Making* (CDM) is defined as a large-scale DM model for leveraging the wisdom of crowds without the need of a consensus process [2]. It captures the smart collective opinions from the natural language evaluations provided by a massive number of people. This kind of free evaluations are analyzed through sentiment analysis (SA) [3], which allows to distill the opinions of the experts from natural language texts.

This paper relies on CDM based on social media opinions, as the natural way to get the crowd opinions in the current digital word with people connected by social media. These platforms are packed with natural language reviews containing the opinions of billions of people scattered throughout the world who constitute a wise group due to their independence, diversity and decentralization. The reviews with the identified opinions can be the input data for CDM, since they constitute a source of wisdom where the experts are the social media users and the alternatives are the entities advertised on the platforms.

Decision-making models are applied in a wide range of disciplines [4–6]. We argue that from a practical point of view these models can be enhanced when they provide some details of their operations. When people ignore why the decision model selects an alternative as the best one, they do not trust on it. We claim that DM methods have to shed more light on the information elements that support their outputs so that people further trust them.

Explainable artificial intelligence (XAI) aims to develop more explainable models while maintaining high accuracy, as well as allowing humans to understand and trust them [7]. There are two closely related significant concepts in XAI [8]: (1) *interpretability*, as the ability to provide meaning in understandable terms to humans; and (2) *explainability*, as an interface between humans and a decision maker that is simultaneously an accurate proxy of the decision maker and comprehensible to humans. Concerning DM, and so CDM, it is suitable to talk about explainability when the mechanism itself is not interpretable but an additional explanatory algorithm is necessary [9]. Despite the research efforts, explainability in this context still lack usability when

[*] Corresponding author.
E-mail addresses: czuheros@ugr.es (C. Zuheros), emcamara@decsai.ugr.es (E. Martínez-Cámara), viedma@decsai.ugr.es (E. Herrera-Viedma), iakatib@kau.edu.sa (I.A. Katib), herrera@decsai.ugr.es (F. Herrera).

https://doi.org/10.1016/j.inffus.2023.101821
Received 6 February 2023; Received in revised form 20 April 2023; Accepted 21 April 2023
Available online 27 April 2023

implemented in practice since users see it as a tool designed for engineers rather than a useful tool for them [10]. Informative and detailed explanations should be included to enhance stakeholders trust the systems.

In this paper, we propose the Explainable Crowd Decision Making based on Subgroup Discovery and Attention Mechanisms (ECDM-SDAM) methodology as an a posteriori explainable process that captures the wisdom of crowds. It is rooted in CDM as it manages unconstrained natural language opinions as expert evaluations posted by social media users. We profit that groups as a whole, such as these users, are more intelligent than an elite few [11]. The ECDM-SDAM is an a posteriori explainable CDM since it has a backward mechanism that provides easy understandable natural language explanations that unveil the relevant information for the decision process, by introducing two key elements:

- Attention mechanisms. The methodology analyzes reviews to elicit the expert opinions using an innovative sentiment analysis method (SAM) based on multi-task deep learning and attention mechanism that we propose and call it Attention based Sentiment Analysis Method (ASAM). The weights of the attention mechanisms stand out the relevant sentences provided by the experts behind the decision reported.
- Subgroup discovery (SD). The methodology involves this technique, which discovers interesting relationship between objects, to identify the relevant aspect terms associated to particular criteria for the decision reported. The opinions that refer to a specific criterion of an alternative are extracted and represented in a novel table that we proposed and call it Bag of Opinions by Criteria (BOC), so we apply SD on them.

The ECDM-SDAM methodology workflow consists in three steps, namely: (1) obtaining the expert opinions, where opinions are extracted by the proposed ASAM model and represented into the proposed BOC tables; (2) crowd decision making, which conducts the DM mechanism itself carrying out a collective aggregation and an exploitation phase; and (3) explainable backward process, that generates explanatory texts that highlight the aspect terms, criteria and sentences that the methodology focused on to reach the solution. The meaningful sentences are identified through the attention mechanism of the ASAM model while the relevant aspect terms are identified applying SD algorithms on the BOC tables.

The main contributions of the proposed ECDM-SDAM methodology are:

- **To provide easy-to-understand explanations in natural language.** It automatically generates plain texts that, on the one hand highlight the strengths on the best alternative, and on the other hand show the weaknesses of the lowest ranked alternative. Both texts are detailed and easy to understand due to their conciseness, clarity and itemized structure.
- **To manage high quality input data.** The methodology handles user opinions as high quality experts evaluations since: (1) they are unconstrained, completely free and can be provided in natural language, which is in line with human reasoning; and (2) they capture the wisdom of the crowd as they come from social media users who are a large group of diverse, independent and decentralized individuals.
- **To highlight the relevant information for the decision.** The methodology itself is informative, pointing out what it has focused on to select an alternative as the best and an alternative as the worst, identifying: (1) its most relevant criteria, *e.g.*, *food* or *drinks*; (2) its most relevant aspect terms associated to a particular criterion, *e.g.*, *rice* or *water*; and (3) its most relevant sentences provided by the experts.

We evaluate the ECDM-SDAM methodology in a real case study. Specifically, it analyzes the restaurants from the TripR-2020Large dataset [2] to identify the best one and to explain the reached solution. This dataset is suitable to evaluate the methodology, since it contains quality evaluations that allows automatically provide explanations understandable to humans.

This paper is structured as follows. Section 2 introduces the basis of our proposal through a review on using opinions for DM, CDM, explainability in DM, attention mechanisms, and SD. Section 3 presents the ECDM-SDAM methodology detailing its three steps. Section 4 solves the case of study from TripR-2020Large by applying the methodology. Section 5 exposes some concluding remarks and future work.

2. Background

This section presents the basic concepts related to the ECDM-SDAM methodology. Section 2.1 exposes the use of social media opinions for decision making. Section 2.2 focuses on crowd decision making and Section 2.3 on explainability in DM. The explanation provided by ECDM-SDAM is mainly achieved by using attention mechanisms (see Section 2.4) and SD (see Section 2.5).

2.1. Using social media opinions for decision making

Social media are crowded of natural language evaluations containing opinions, which can be used as input data for CDM. These opinions are very significant and worthwhile to process given the intense use of such platforms by billions of users to express their experiences. This is a natural environment to get the wisdom of crowds in the current digital word. In this context, an opinion is given by [12]: (1) an aspect term or aspect as the target of opinion, (2) a category as a group of similar aspects, and (3) a polarity as the sentiment value.

TripAdvisor is a well known social media website where users share their opinions on travels. Zuheros et al. [2] created and released the TripR-2020Large[1] dataset which contains real reviews from restaurants advertised on TripAdvisor with manually annotated opinions. The reviews are natural language texts written in English, optionally complemented by numerical ratings associated to particular criterion. The texts are divided into sentences and each sentence presents all the user opinions at the aspect level. It contains thousands of opinions that can be used for evaluating CDM models. Thus, we will evaluate our methodology into the TripR-2020Large dataset.

An example of a brief real natural language evaluation from the TripR-2020Large dataset is given by: *"Classic and always a great atmosphere. Amazing wine list also"*. It consists of two sentences and each of them presents an opinion. The first sentence presents a *positive* opinion on the aspect *atmosphere* regarding the category *ambience* of a restaurant, so the first opinion can be represented by (*atmosphere*, *ambience*, *positive*). The second sentence presents a *positive* opinion on the aspect *wine list* regarding the category *drinks*, so the second opinion can be represented by (*wine list*, *drinks*, *positive*). The evaluations can be analyzed by a SAM to extract the opinions, which feed a CDM model where criteria correspond to the extracted opinion categories.

There are several studies that take advantage of opinions and social media to create decision models. In the following, we analyze shortly them: (1) Punetha et al. [13] present an unsupervised SA system that makes it easier for customers to make purchasing decisions based on emotions of reviews, by combining multi criteria DM and game theory, (2) Tayal et al. [14] propose a multi criteria DM for multi-aspect based personalized ranking of the products from e-commerce websites taking into account the customer preferences, (3) Zhu et al. [15] develop a method for dynamic collaboration between public and experts in large-scale group emergency DM via text data from social media, (4)

[1] https://github.com/ari-dasci/OD-TripR-2020Large

Morente-Molinera et al. [16,17] extract valuable information from debate texts taking place on social media using DM and lexicon-based approaches, (5) Zuheros et al. propose in [12] a DM methodology that considers TripAdvisor natural language reviews as expert evaluations where opinions are extracted by a semantic understanding process, and in [2] a DM model that captures the wisdom of the crowd that is naturally offered on this social media, and (6) finally, there exist a prominent stream of studies emerging in this field that show its value [18–20]. These studies are hardly comparable since they are highly diverse, using different types of input data, analyzing opinions at different levels of processing, lacking in explainability, and in some cases not performing opinion processing per se.

2.2. Crowd decision making

The *wisdom of crowds* [11] states that groups, under the right circumstances, are remarkably intelligent and are often smarter than the smartest people in them. Even if many people in the group are not especially well-informed, a collectively robust decision can be reached. The necessary conditions for the crowd to be wise are the independence, the diversity, and the decentralization of the group. Thus, the forced interaction between group members is inappropriate.

Our hypothesis is that the natural way to capture the wisdom of the crowd, in the current scenario of the digital society, is to extract the opinions that are published on social media by means of natural language processing. The opinions can be the input of a DM system so that it harness the wisdom of crowds. There exists some studies that already connect such fields: the wisdom of the crowd, DM, and social media. Herrera-Viedma et al. [21] present the challenges of social networks for DM frameworks guided by the wisdom of the crowd. Verasius et al. [22] plan to analyze social media to better capture the wisdom of crowds for their proposed tourism chatbot, which returns the preferred sites ranked based on the judgment of the crowds from questionnaires.

A novel approach in DM integrates the wisdom of crowds to better capture the natural collective intelligence, leading to the Crowd Decision Making (CDM) [2]. The concept of CDM refers to DM models that leverage the wisdom of the crowd by integrating unconstrained evaluations of a large amount of people. It captures the natural knowledge of a wise group of individuals since it avoids the consensus processes and handles free natural language texts by incorporating SA tools. Social media platforms are perfect environments for CDM, since they allow to evaluate varied entities through natural language reviews and to connect people around the world who are independent, diverse and decentralized.

2.3. Explainability in decision making

Explainability can be understood as an interface between humans and a decision maker that is at the same time both an accurate proxy of the model and comprehensible to humans [7]. The goal of including it in DM models is to underpin their sense-making procedure to make relevant information available and understandable, making wider their applicability in real environments.

In the literature, we find some initial studies that try to incorporate explainability into DM models [23,24]. For example, Toni et al. [25] provide explanations to understand the underpinning mechanisms of DM models using dispute trees by means of computation argumentation. However, we cannot find DM models that manage opinions from natural language evaluations and that incorporate explainability mechanisms to justify their achieved ranking of alternatives.

How to fix the idea of explainability in DM is still a challenge as it is discussed in [26] and there are some difficulties to overcome: (1) to provide explanations easily understandable in natural language; (2) to highlight the reasons behind the best alternative instead of making pairwise comparisons between the best option and the rest of alternatives; and (3) to manage free expert evaluations.

2.4. Attention mechanisms in natural language processing

Natural language processing (NLP) is the interdisciplinary field of artificial intelligence and linguistic for understanding and generating human language. The attention mechanism was first introduced in NLP by Bahdanau et al. [27] as an extension to the encoder–decoder model to cope better the dependencies on long sentences avoiding to squash all the information. The main idea is to generate a weight distribution associated to the input sentence so that relevant words have higher values [28].

A wide range of attention mechanisms have been proposed for tackling NLP tasks. This work focuses on SA, the NLP task concerned with determining the author's attitudes toward an object or the general emotional tendencies of texts [3]. Therefore, we establish two categories that involve attention mechanism in NLP based on the SA task: (1) methods for sentence encoding [29] and (2) methods for SA [30,31]. For example, Huang et al. [32] provide sentence representations by means of self-attention methods based on different window sizes. Regarding SA, Zhang et al. [3] recognize the general attitude of texts by including transformers into a multitask network.

There is much discussion about whether attention mechanisms offer the explanation of neural network models [33,34]. We agree with Feng et al. [23] that attention scores provide plausible rationales for their use at practical level, even though they may not provide a complete internal justification for the behavior of the model.

2.5. Subgroup discovery

Subgroup discovery (SD) is a widely used technique focused on discovering interesting relationships between objects in a dataset with respect to an specific property of interest, to which they should have the most unusual statistical characteristics [35]. The extracted patterns consists of induced subgroup descriptions known as rules, which take the form *antecedent* → *consequent* so that the *antecedent* is a set of features and the *consequent* is a value of the target variable of interest.

The metrics used in SD rely on association rules [36]. Consider a rule R: $X \rightarrow Y$, the number of total instances N of a dataset, and a function $n(x)$ to count the frequency of x in the dataset. The most common evaluation metrics are: (1) the support, that measures the frequency of rules containing both X and Y by $Support(R) = n(X \cdot Y)/N$, and (2) the confidence, that provides the likelihood of Y appearing in those rules that contain X by $Confidence(R) = n(X \cdot Y)/n(X)$. Additionally, a metric to measure the level of unusualness of a rule is the normalized weighted relative accuracy (NWRAcc) which is computed by

$$NWRAcc(R) = \frac{WRAcc(R) - LB(R)}{UB(R) - LB(R)}$$

where $WRAcc(R) = nCon \cdot [Confidence(R) - nAnt]$, $nCon = n(Y)/N$, $nAnt = n(X)/N$, $UB(R) = (1 - nAnt) \cdot nAnt$ and $LB(R) = (1 - nAnt) \cdot (-nAnt)$. NWRAcc values greater than 0.5 mean there is a good level of unusualness.

3. Explainable crowd decision making methodology

We propose the Explainable Crowd Decision Making based on Subgroup Discovery and Attention Mechanisms (ECDM-SDAM) methodology as a CDM model [2] able to explain the result it reaches. It captures the wisdom of the crowd available on social networks from natural language expert evaluations and provides an explanation that justifies one alternative being chosen as the best by identifying its strongest points. Additionally, it offers a negative explanation for the worst alternative to identify its weakest points. The methodology is composed of three steps namely, *obtaining the expert opinions* (see Section 3.1), *crowd decision making* (see Section 3.2), and *explainable backward process* (see Section 3.3). Fig. 1 depicts the steps of the methodology.

The ECDM-SDAM methodology obtains the final evaluation after analyzing the evaluations by extracting their most relevant information.

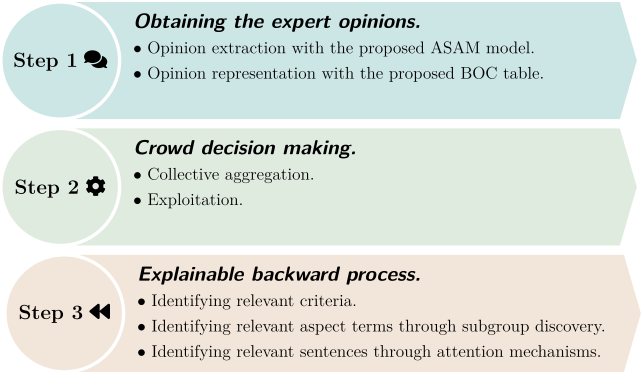
Step 1
Obtaining the expert opinions.
• Opinion extraction with the proposed ASAM model.
• Opinion representation with the proposed BOC table.
Step 2
Crowd decision making.
• Collective aggregation.
• Exploitation.
Step 3
Explainable backward process.
• Identifying relevant criteria.
• Identifying relevant aspect terms through subgroup discovery.
• Identifying relevant sentences through attention mechanisms.

Fig. 1. Workflow of the ECDM-SDAM methodology.

It considers user reviews from social media platforms as expert evaluations of a DM problem. These reviews are analyzed by a sentiment analysis method based on an attention mechanism that we call ASAM, which infers the opinions of the users at aspect level. We represent the opinions into a table structure that we call BOC. Then, the crowd decision making process is performed by conducting a collective evaluation that aggregates the opinions of all users and by carrying out an exploitation process that obtains the classification of the alternatives. The last step of the methodology conducts a backward process to explain the reached solution. It identifies the relevant information for the decision process to select an alternative as the best and generates explanatory texts in natural language. The explanation generation process is carried out through three stages, which are integrated into the final evaluation: (1) the most relevant criteria are extracted taking into account the collective evaluation and the importance of the criteria for the users, (2) the most relevant aspect terms are extracted by identifying in the BOC structures the opinion objects with the greatest positive and negative influence for each alternative and criterion by applying a subgroup discovery algorithm, and (3) the most relevant sentences are extracted by selecting the utterances from the expert text evaluations that have the highest weights in the attention mechanism of the ASAM neural network.

3.1. Obtaining the expert opinions

The objective of the ECDM-SDAM methodology is to rate a set of alternatives $X = \{x_1, \dots, x_n\}$, which are evaluated by a massive number of users of a social network platform. These users are considered as the set of experts $E = \{e_1, \dots, e_l\}$. We crawl all their reviews about the alternatives, which are written in plain text and optionally supplemented with numerical ratings, provided on the website. We collect the text provided by the expert e_k, $k = 1, \dots, l$ for evaluating the alternative x_i, $i = 1, \dots, n$ into the element t_i^k. There may be lack of information as some experts may not evaluate all the alternatives, so some element t_i^k may not exist. This stage analyzes the texts to infer the expert opinions. In Section 3.1.1, we explain how we extract the opinions expressed in t_i^k by means of the novel ASAM model, and we describe in Section 3.1.2 how we represent such opinions into the novel BOC table.

3.1.1. Opinion extraction with the ASAM model

The natural language texts t_i^k contain a great amount of valuable information about what the user e_k thinks about the alternative x_i. The ECDM-SDAM methodology infers the expert opinions from these elements at aspect level, so it actually conducts an aspect-based sentiment analysis (ABSA) task. Particularly, we build a SAM based on attention mechanisms to extract the opinions, that we call Attention based Sentiment Analysis Method (ASAM). It is a multi-task neural network model, inspired by DOC-ABSADeepL [12], that incorporates a novel attention mechanism. Fig. 2 presents its architecture and, subsequently, we describe its main components:

Input layer. The model analyzes s-length reviews $\{w_1, \dots, w_s\}$. The words w_r, $r = 1, \dots, s$ are represented by word embeddings so that each word is a d-dimensional vector $\mathbf{we}_r = (we_{r1}, \dots, we_{rd})$.

Multitask-learning. The model analyzes the input sentences to infer the opinions by means of two blocks of information processing:

- The aspect and category processing block. It is responsible of inferring the aspect terms, which we call them aspects, and its associated category. We apply a Bidirectional Gated Recurrent Unit (BiGRU) and the novel attention mechanism that we subsequently describe.
- The polarity processing block. It is responsible of extracting the sentiment values. We apply a Bidirectional Long Short Term Memory (BiLSTM), the novel attention mechanism, a fully-connected layer, and a dropout layer to prevent over-fitting.

The attention mechanism applied in both blocks aims at dynamically pointing out the relevant features of the input sentence. The mechanism is based on the scaled dot-product attention proposed by Vaswani [37]. It considers as input the output of the BiGRU or BiLSTM which is a matrix $X = [\mathbf{x}_1, \dots, \mathbf{x}_s] \in \mathbb{R}^{s \times 2h_{rnn}}$ where h_{rnn} is its number of units. The attention mechanism outputs a vector of weights α as follows:

$$\alpha = \left(- \left| \frac{qX^T}{\sqrt{d_X}} \right| \right)^T \tag{1}$$

where $q \in \mathbb{R}^{1 \times 2h_{rnn}}$ is a parameter learning vector initialized by an uniform distribution and d_X is the dimension of X. The vector of weights α manifests the attention assigned to each word of the input sentence. Finally, we compute the attention matrix by multiplying α and the corresponding BiLSTM or BiGRU output as follows:

$$M = \alpha X, M \in \mathbb{R}^{s \times 2h_{rnn}}. \tag{2}$$

The attention mechanism allows the network to learn the relevant words. When predicting aspects and categories, the model should focus on one type of words while predicting polarities should focus on another sort. For example, nouns usually correspond to terms and entities while adjectives and adverbs tend to express sentiments. The ASAM model captures this phenomenon by including an attention layer for each block separately.

Output layer. The model has three classification layers so that each one provides a component of an opinion: (1) the *aspect layer* points out whether the input word is an opinion aspect term; (2) the *category layer* provides the category to which the aspect term belongs in case the input word expresses an opinion; and (3) the *polarity layer* indicates a *positive*, *negative* or *neutral* sentiment about the identified aspect term in case the user has expressed an opinion.

We represent the opinions extracted by the ASAM model as a vector $(aspect, category, polarity)$, where $aspect$ is the term of interest identified by the aspect layer, $category$ is the category to which the aspect term belongs identified by the category layer, and $polarity$ is the sentiment value of the opinion identified by the polarity layer. ASAM analyzes the natural language text t_i^k provided by the expert e_k to the alternative x_i delivering q_i^k opinions. We define the set of opinions associated to the expert e_k and the alternative x_i inferred from t_i^k by:

$$Opinions_i^k = \{op(t_i^k)_1, \dots, op(t_i^k)_{q_i^k}\}, q_i^k \in \mathbb{N} \tag{3}$$

where $op(t_i^k)_{q_i^k} = (aspect(t_i^k)_{q_i^k}, category(t_i^k)_{q_i^k}, polarity(t_i^k)_{q_i^k})$ is the q_i^k-nth opinion of the expert e_k for the alternative x_i.

We consider the categories identified by the ASAM model as criteria that we join to the criteria that experts evaluate directly on the website through numerical ratings. The set of criteria is compiled into $C = \{c_1, \dots, c_m\}$.

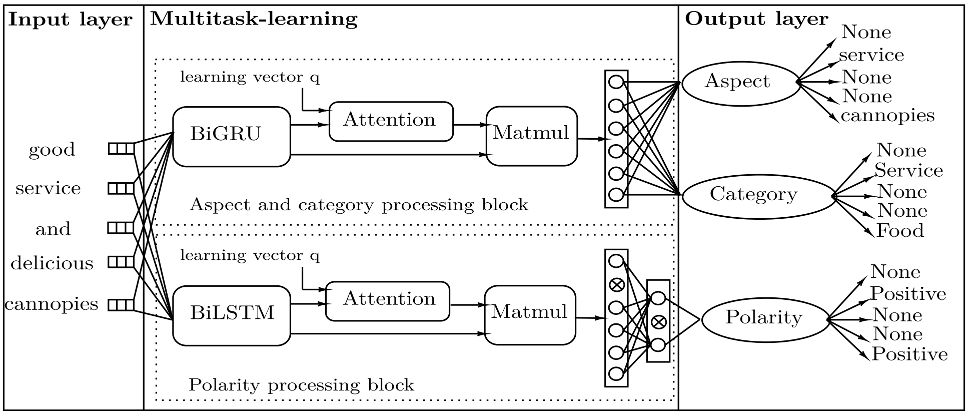

Fig. 2. Architecture of the ASAM model for extracting the expert opinions.

3.1.2. Opinion representation with the BOC table

Once the opinions are extracted from the natural language texts t_i^k, we represent them into a table that we call bag of opinions by criteria or BOC. It captures all the opinions provided by the experts about an alternative based on a particular criterion. First, we collect the opinions referring to the criterion c_j for the alternative x_i by:

$$Opinions_i(c_j) = \{op_1, \dots, op_g\} = \bigcup_{k=1}^{l} Opinions_i^k(c_j) \tag{4}$$

where $Opinions_i^k(c_j) = \{op(t_i^k)_1, \dots, op(t_i^k)_{q_i^{k'}}\}$, $q_i^{k'} \leq q_i^k$ such that $op(t_i^k)_h = (aspect(t_i^k)_h,\ c_j,\ polarity(t_i^k)_h)$, $h = 1, \dots, q_i^{k'}$ compiles the opinions provided by the expert e_k for evaluating the alternative x_i on the criterion c_j. When no expert evaluates c_j for x_i, then $Opinions_i(c_j) = \{\emptyset\}$.

The opinions collected into $Opinions_i(c_j)$ refers to the criterion c_j of the alternative x_i. These opinions may evaluate different aspects and entail different opinion polarity meanings. We collect all the aspects that appear in at least one opinion from the non-empty set $Opinions_i(c_j) = \{op_1,\ \dots,\ op_g\}$ into a set of aspects $Aspects_i(c_j) = \{asp_i(c_j)_1,\ \dots\ ,asp_i(c_j)_{g'}\}$, $g' \leq g$. Furthermore, such opinions are associated to particular sentences from t_i^k, that we collect into the set of sentences $Sentences_i(c_j) = \{sent_1,\ \dots,\ sent_{g''}\}$, $g'' \leq g$. We define the BOC table to capture the aspect terms and polarities of the opinions from the sentences that evaluate the criterion c_j of the alternative x_i as follows:

Definition 1 (*Bag of Opinions by Criteria (BOC)*)**.** Let $Opinions_i(c_j) = \{op_1,\dots,op_g\}$ be a non-empty set collecting all the expert opinions provided to the alternative x_i for the c_j criterion, let $Aspects_i(c_j) = \{asp_i(c_j)_1,\ \dots,\ asp_i(c_j)_{g'}\}$, $g' \leq g$ be its associated set of aspect terms, and let $Sentences_i(c_j) = \{sent_1,\ \dots,\ sent_{g''}\}$, $g'' \leq g$ be its associated set of sentences. We define a bag of opinions by criteria (BOC) as a table that captures the opinions of each sentence as rows of vectors. It presents each opinion op_u, $u = 1, \dots, g$, as an array $(sent_u, asp_i(c_j)_{u1}, asp_i(c_j)_{u2},\ \dots,\ asp_i(c_j)_{ug'}, pol_u)$ such that $sent_u$ is the sentence to which op_u belongs, $asp_i(c_j)_{uv} \in \{0,1\}$ $\forall v = 1, \dots, g'$ are binary features representing the presence or absence of the aspect $asp_i(c_j)_v$ in op_u, and pol_u indicates whether the opinion sentiment is *negative* or *positive*. We merge two rows by summing the aspect terms elements when they have the same sentence and polarity elements.

Table 1 shows an abstract representation of a BOC. It collects the opinions provided by all the experts for each alternative regarding each criterion c_j, $j = 1, \dots, m$, so we build m tables for each alternative. We represent the BOC associated to the alternative x_i concerning the criterion c_j as $BOC_i(c_j)$. If no expert evaluates criterion c_j for the alternative x_i, then $BOC_i(c_j) = \{\emptyset\}$.

Table 1
Abstract representation of a BOC table, which compiles the opinions of all the experts referring a criterion of an alternative.

Sentence	asp$_1$	asp$_2$	...	asp$_{g'}$	Polarity
$sent_1$ with one positive opinion to asp_2	0	1	...	0	positive
$sent_2$ with one positive opinion to asp_1	1	0	...	0	positive
$sent_2$ (again) with one negative opinion to asp_2	0	1	...	0	negative
...	...	...	...	...	...
$sent_{g''}$ with positive opinions to asp_1 and $asp_{g'}$	1	0	...	1	positive

3.2. Crowd decision making

Experts evaluate the alternatives through reviews written in plain text optionally complemented by numerical ratings. The previous step extracts the expert opinions from the texts. At this moment, the ECDM-SDAM methodology focuses on selecting the best alternative through a CDM model. It leverages the wisdom of the crowds from a large number of people without a consensus process following the traditional scheme of DM models, which are composed of two main phases: (1) an aggregation step to combine all the expert evaluations into a collective evaluation (see Section 3.2.1) and (2) an exploitation step to rank the alternatives providing the ranking solution (see Section 3.2.2).

3.2.1. Collective aggregation

The opinions of all the expert evaluations are represented into the BOC tables. Specifically, for each alternative x_i, $i = 1, \dots, n$, there are m tables $BOC_i(c_j)$, one per criterion c_j, $j = 1, \dots, m$, collecting the expert opinions. According to Definition 1, each table has associated a particular set of aspect terms $Aspects_i(c_j) = \{asp_i(c_j)_1,\ \dots,\ asp_i(c_j)_{g'}\}$. The collective aggregation of the experts measures the balance between the positive and the negative opinions by focusing on the aspects and the polarity component of the BOC tables. Particularly, we obtain the collective evaluation by taking into account the frequency of aspects associated with a positive sentiment among the total of opinions. Hence, we build a collective textual evaluation (CTE) $n \times m$ matrix such that each element $cte_{i,j}$, $i = 1, \dots, n$, $j = 1, \dots, m$ is computed by aggregating all the elements of the non-empty $BOC_i(c_j)$ table as follows:

$$cte_{i,j} = \frac{\sum_{p=1}^{g'} Freq(asp_i(c_j)_p, positive)}{\sum_{p=1}^{g'} Freq(asp_i(c_j)_p)} \in [0,1] \tag{5}$$

where $Freq(asp_i(c_j)_p)$ is the amount of elements of the BOC table in which the aspect $asp_i(c_j)_p$, $p = 1, \dots, g'$, is present with wherever polarity for the alternative x_i and the criterion c_j, and $Freq(asp_i(c_j)_p, positive)$

is associated to positive polarity. When $BOC_i(c_j) = \{\emptyset\}$, then $cte_{i,j}$ is a not available value. The $cte_{i,j}$ values higher than 0.5, respectively lower, represent there is a predominant collective positive sentiment, respectively negative, for x_i about c_j.

Optionally, experts provide numerical ratings to evaluate the alternatives. We collect them into an individual numerical evaluation (INE^k) matrix for each expert e_k, $k = 1, \ldots, l$. Traditionally, these values are provided in a five point likert scale. When an expert evaluates the same criterion of an alternative several times through numerical ratings, we average them. Then, we aggregate the INE matrices to get a collective numerical evaluation (CNE) $n \times m$ matrix by averaging the existing ine values and transforming them to the interval $[0, 1]$:

$$cne_{i,j} = \frac{\sum_{k=1}^{l'} ine^k_{i,j} - l'}{4l'} \in [0, 1], \tag{6}$$

where $l' \leq l$ is the amount of experts that evaluate the alternative x_i on criterion c_j through numerical ratings. When $l' = 0$, then $cne_{i,j}$ is a not available value.

Finally, we build the collective evaluation (CE) matrix which compiles the textual and numerical evaluations of all the experts for the alternative x_i related to criterion c_j. Thus, it aggregates the CTE and the CNE matrices by:

$$ce_{i,j} = \frac{cte_{i,j} + cne_{i,j}}{2} \in [0, 1]. \tag{7}$$

We capture the missing information as follows. If the $cte_{i,j}$ value is not available, then $ce_{i,j} = cne_{i,j}$, and if the $cne_{i,j}$ value is not available, then $ce_{i,j} = cte_{i,j}$.

3.2.2. Exploitation

It obtains the final evaluation of each alternative by aggregating the evaluations of all its criteria, and orders them giving rise to the ranking of alternatives. We weight the criteria according to their importance for the experts [2]. Particularly, we compute the weight ω_j associated to the criterion c_j, $j = 1, \ldots, m$, through the relative frequency of $n\_evaluations(c_j)$, that is the sum of opinions and numerical ratings provided to c_j, among the total of evaluations n_total by:

$$\omega_j = \frac{n\_evaluations(c_j)}{n\_total}. \tag{8}$$

The evaluation for each alternative x_i, $i = 1, \ldots, n$ is computed by means of a weighted average of the collective evaluation associated with the criteria by the criteria weight. Then, we build the final evaluation (**fe**) vector by:

$$fe_i = \sum_{j=1}^{m} (\omega_j \times ce_{i,j}) \in [0, 1] \tag{9}$$

where ω_j is the weight associated to the criterion c_j obtained by Eq. (8). We obtain the final ranking of the alternatives that solves the DM problem ordering thefe_i values. The first element of the ranking is the best alternative.

3.3. Explainable backward process

The ECDM-SDAM methodology is an a posteriori explainable process because it is based on a backward mechanism that allows to partially explain its final solution. It focuses on the final ranking and provides an explanatory text justifying why certain alternative is chosen as the best in the ranking, which we named x_{best}. This provides a positive explanation for the best alternative that focuses on identifying its strongest or most beneficial points. Additionally, the ECDM-SDAM methodology provides a negative explanatory text for the worst alternative in the ranking, that we named x_{worst}, in order to identify its main points to improve. Both explanatory texts, the positive for x_{best} and the negative for x_{worst}, shed light on identifying the criteria, the aspect terms, and the sentences that have the greatest impact on their evaluations. Subsequently, we describe the three stages of the explainable backward process that provide the three elements identified into the texts:

- **Stage 1: Identifying relevant criteria.** It extracts relevant information associated to criteria by analyzing the collective evaluations (see Section 3.2.1). The CE matrix provides the collective evaluation and their values can be interpreted as $[0, 1]$ ratings that measure how good is a criterion for an alternative considering all the expert evaluations. A criterion may achieve a perfect score of one into the CE matrix, but it may contribute very little to the final score of the alternative. It is necessary to weight up the collective evaluation of each criterion with its importance to understand its impact on each alternative. Thus, we build a weighted collective evaluation (WCE) matrix such as the collective evaluation is weighted on criteria (see Eq. (8)) by:

$$wce_{i,j} = \omega_j \times ce_{i,j}. \tag{10}$$

This stage identifies the most beneficial criterion for the best alternative x_{best} and the most detrimental criterion for the last alternative x_{worst} based on the CE and the WCE matrices. Regarding x_{best}, it provides the fragment of the positive explanatory text that identifies the most beneficial criterion as the one with the highest value into the WCE matrix by:

$$BestCriterion = argmax_j(wce_{x_{best},j}). \tag{11}$$

We focus on that criterion for x_{best} and provide the collective evaluation as a rating score in the interval $[0, 10]$, since it is more natural for humans, by computing:

$$Rating = 100 \times ce_{x_{best},BestCriterion}. \tag{12}$$

It provides the following item from the explanatory text: *"Its criterion of greatest interest, [$BestCriterion$], reaches a rating of [$Rating$] out of 10"*.
Regarding the negative explanatory text, it identifies the most detrimental criterion as the one with the lowest value into the WCE matrix for x_{worst} by $WorstCriterion = argmin_j(wce_{x_{worst},j})$. It provides the extract of the explanatory text: *"Its most detrimental criterion is [$WorstCriterion$]"*.

- **Stage 2: Identifying relevant aspect terms through subgroup discovery.** The methodology already identifies the most significant criterion associated with x_{best} and x_{worst}. This stage identifies the aspect terms associated with these criteria that have the greatest impact for both alternatives. We handle it by applying SD techniques into the BOC tables (see Section 3.1.2), as they identify the evaluated aspect terms for each criterion and alternative. The SD method finds the population subgroups, which are aspect terms of the opinions provided by the experts, that are statistically interesting with respect to a property of interest, that is the polarity of the opinions. We consider two sorts of subgroups to discover based on the target of interest: aspect terms with positive polarity for x_{best} and aspect terms with negative polarity for x_{worst}.
We apply the Apriori-SD algorithm [38] since it is a classic competitive SD method for working with binary categorical features, such as the elements of the BOC tables. Furthermore, in contrast to traditional SD methods, it provides smaller rule sets for better understanding and the individual rules have higher coverage and significance. The main peculiarity of Apriori-SD relies on the post-processing rule subset selection where it considers the WRAcc metric to evaluate the quality of the induced rules. In our proposal, we adapt such produce to sort rules from best to worst based on the NWRAcc metric, then it decreases the weight of the covered samples by the best rules, and repeats the process until all the samples from the BOC tables are covered or there are no more rules which are previously produced by the apriori algorithm with a minimal support and confidence.
Regarding the positive explanatory text, this stage generates association rules from the opinions provided to the best alternative regarding its best criterion, *i.e*, from $BOC_{x_{best},BestCriterion}$,

whose *consequent* is *positive*. We compile the *antecedent* aspect terms with highest support and named it $PositiveAspects$. It provides the following extract of the explanatory text: *"The [list of $PositiveAspects$] stand out positively".*

Regarding the negative explanatory text, we perform a similar procedure on the BOC_{x_{worst},c_j}, $\forall j = 1, \ldots m$ tables collecting all the aspects terms to improve into $NegativeAspects$, and it provides the following extract of the explanatory text: *"The [list of $NegativeAspects$] stand out negatively".*

- **Stage 3: Identifying relevant sentences through attention mechanisms.** We analyze the ASAM model (see Section 3.1.1) to identify the sentences with higher impact on the decision process of the methodology. Its attention mechanism assign two weights to each input word: one associated to the aspect and category processing block, and another associated to the polarity processing block. Independently for both types of weights, we normalized them so that all the reviews have the same importance. We average both types of weights to get the final attention weight for each word, as it results in a better quality solution than considering only one weight. Let $Sentence_{s'} = \omega_1, \ldots, \omega_{s'}$, $s' \leq s$ be a sentence of a s-length review, such as each word ω_v, $v = 1, \ldots, s'$ has associated a final attention weight $Attention(\omega_v)$. We assign the sentence attention weight by:

$$Attention(Sentence_{s'}) = \frac{\sum_{v=1}^{s'} Attention(\omega_v)}{s'}. \qquad (13)$$

We build sentence rankings to compile the succinct meaningful sentences based on their attention scores. These rankings can be generated based on different filters that can be combined: (1) by the alternative that evaluates the sentence, (2) by the sentiment expressed by the sentence, and (3) by the criterion evaluated by the sentence. This way, we can build a ranking with the most important sentences that evaluate positively certain alternative based on a particular criterion. Furthermore, we can combine multiple rankings. We denote by *Ranking* the ranking of interest. This stage provides the following extract for the positive explanatory text: *"[Number of most significant sentences to show] of the expert evaluations that most benefit this alternative [to change by the topic of interest] being selected as the best are: [list of Ranking]".*

This stage provides the following extract for the negative explanatory text: *"[Number of most significant sentences to show] of the expert evaluations that most harm this alternative [to change by the topic of interest] being selected as the last one are: [list of Ranking]".*

4. Case of study: choice of a restaurant

This section presents the use of the ECDM-SDAM methodology in a real case of study for choosing restaurants. Section 4.1 obtains the expert opinions. Section 4.2 develops the CDM model to get the ranking of the restaurants. Section 4.3 presents the explainable backward process to understand the reached solution.

4.1. Obtaining the expert opinions

We use the TripR-2020Large[2] dataset to evaluate the ECDM-SDAM methodology. The TripR-2020Large dataset is composed of reviews from 4 London restaurants posted by 132 users through the TripAdvisor platform [2]. The set of alternatives is $X = \{x_1, x_2, x_3, x_4\} = \{$*Oxo Tower Restaurant, J. Sheekey, The Wolseley, The Ivy*$\}$ and the set of experts is $E = \{e_1, \ldots, e_{132}\}$. They provide 474 natural language reviews composed of 2,522 sentences and optionally supplemented by numerical ratings in a five point likert scale.

We process the opinions of the reviews with the ASAM model (see Section 3.1.1). It is trained on the training set of the SemEval-2016 dataset,[3] since it is a widely used aspect based sentiment analysis dataset on the restaurant reviews domain [39]. We set $s = 200$ input words since most of the sentences of this dataset are less than 200 words. We use the $d = 300$ dimensional FastText word embeddings trained on Common Crawl.[4] The remaining hyper-parameters are $h_{rnn} = 128$ for both LSTM and GRU layers, $dropout = 0.1$, $batch\_size = 4$, and $epochs = 50$ with earlystopping. Table 2 shows the results of the ASAM model tested on the testing set of SemEval-2016 and on the TripR-2020Large dataset. We compare it with the results obtained by the DOC-ABSADeepL model [2] and conclude that the proposed model, called ASAM, better identifies all the components of the opinions.

We collect the categories identified by ASAM into the set of criteria $C = \{c_1, c_2, c_3, c_4, c_5, c_6\} = \{$*restaurant, food, service, ambience, drinks, location*$\}$. TripAdvisor allows users to rate the first three criteria using numerical ratings.

The BOC tables collect the extracted opinions (see Section 3.1.2). We build 6 tables, one per criterion, associated to each restaurant, resulting in 24 BOC tables. They have a wide variety of aspects associated and at least one sentence with one opinion, so there are no empty tables. For example, the BOC table associated to the x_4 restaurant and the $food$ criterion collects 181 sentences related to 144 aspects. Fig. 3 shows excerpts of the reviews provided by experts who evaluate x_4, the opinions extracted by ASAM, and part of the $BOC_4(food)$ table. Note that the first sentence is not included in the BOC table since the extracted opinion refers to the *service* criterion instead of *food.*

4.2. Crowd decision making

The $BOC_i(c_j)$ table collects the opinions of all the experts provided to the restaurant x_i, $i = 1, \ldots, n$ referring to criterion c_j, $j = 1, \ldots, m$. We aggregate all their elements by means of Eq. (5) to obtain the collective textual evaluation for this criterion and restaurant. For example, the cte value associated to the BOC table from Fig. 3, which is a reduction of the actual $BOC_4(food)$ table due to its large actual dimension, is $cte = \frac{1+1+1+1}{1+1+2+1} = \frac{4}{5} = 0.8$. We perform this process for all the BOC tables getting the CTE matrix.

The collective aggregation also takes into account the optional numerical ratings provided by the experts to certain criteria. We aggregate the numerical values provided to the restaurant x_i and criterion c_j through Eq. (6) getting the CNE matrix. For example, the cne value associated to the numerical ratings provided by the three experts from Fig. 3 is $cne = \frac{3+5-2}{4\times 2} = 0.75$. We aggregate the CTE and the CNE matrices through Eq. (7) getting the CE matrix, which is shown in Fig. 4. It manifests that location is an absolute positive criterion for all the considered restaurants and all of them are high quality.

The exploitation step analyzes the CE matrix to generate the final ranking by weighting the criteria based on their importance for the experts. We compute the criteria weights by means of Eq. (8) getting $w_1 = 0.3$, $w_2 = 0.34$, $w_3 = 0.22$, $w_4 = 0.03$, $w_5 = 0.08$ and $w_5 = 0.03$. For example, the weight associated with the *food* criterion is $w_2 = \frac{1{,}073}{3{,}172} = 0.34$ since experts provide 3,172 evaluations, of which 1,073 refer to *food*. The ECDM-SDAM methodology obtains the final evaluation value for each restaurant by means of Eq. (9) getting $fe_1 = 0.868$, $fe_2 = 0.91$, $fe_3 = 0.881$ and $fe_4 = 0.912$. Thus, the final ranking is $x_4 > x_2 > x_3 > x_1$, *i.e., The Ivy>J. Sheekey>The Wolseley>The Oxo Tower.*

[2] https://github.com/ari-dasci/OD-TripR-2020Large

[3] http://alt.qcri.org/semeval2016/task5/

[4] https://fasttext.cc/docs/en/english-vectors.html

Table 2
Quality of the ASAM model trained on the training set of the SemEval-2016 dataset and tested on the test set of the SemEval-2016 and the TripR2020-Large datasets.

	SemEval-2016			TripR-2020Large		
	Aspect (%F1)	Category (%F1)	Polarity (%Acc)	Aspect (%F1)	Category (%F1)	Polarity (%Acc)
DOC-ABSADeepL	72.5	69.74	72.72	66.7	68.84	72.02
ASAM	**80.27**	**71.8**	**74.4**	**75**	**69.96**	**73.37**

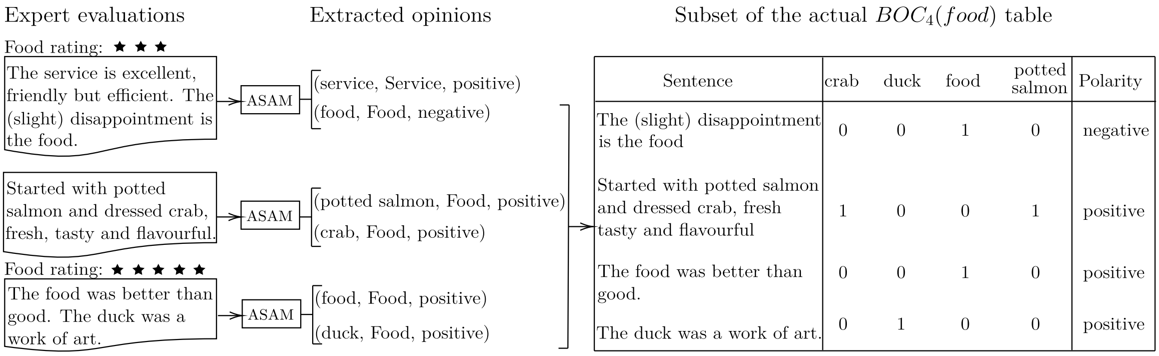
Expert evaluations
Extracted opinions
Subset of the actual $BOC_4(food)$ table
Food rating: ★ ★ ★
The service is excellent, friendly but efficient. The (slight) disappointment is the food.
ASAM
(service, Service, positive)
(food, Food, negative)
Started with potted salmon and dressed crab, fresh, tasty and flavourful.
ASAM
(potted salmon, Food, positive)
(crab, Food, positive)
Food rating: ★ ★ ★ ★ ★
The food was better than good. The duck was a work of art.
ASAM
(food, Food, positive)
(duck, Food, positive)
Sentence
crab
duck
food
potted salmon
Polarity
The (slight) disappointment is the food
0
0
1
0
negative
Started with potted salmon and dressed crab, fresh tasty and flavourful
1
0
0
1
positive
The food was better than good.
0
0
1
0
positive
The duck was a work of art.
0
1
0
0
positive

Fig. 3. Opinions extracted from 3 reviews evaluating the x_4 restaurant by the ASAM model, and representation of the opinions associated to the *food* criterion using the BOC table.

$$\begin{array}{l c} & \begin{array}{cccccc} \text{restaurant} & \text{food} & \text{service} & \text{drinks} & \text{ambience} & \text{location} \end{array} \\ \begin{array}{l} \text{Oxo Tower} \\ \text{J. Sheekey} \\ \text{The Wolseley} \\ \text{The Ivy} \end{array} & \begin{bmatrix} 0.88 & 0.84 & 0.84 & 0.97 & 0.96 & 1 \\ 0.91 & 0.91 & 0.86 & 0.95 & 0.97 & 1 \\ 0.9 & 0.86 & 0.86 & 0.95 & 0.89 & 1 \\ 0.92 & 0.89 & 0.883 & 1 & 0.97 & 1 \end{bmatrix} \end{array}$$

Fig. 4. Collective Evaluation matrix.

4.3. Explainable backward process

The ECDM-SDAM methodology conducts a backward process to automatically point out the relevant information of its sense-making procedure. It provides a positive explanatory text to justify why the restaurant *The Ivy* is the best in the ranking and a negative explanatory text to justify why the restaurant *The Oxo Tower* is the worst one. In the following, we describe how both texts are generated through the three stages of the backward process.

- **Stage 1: Identifying relevant criteria.** We build the WCE matrix (see Fig. 5), as explained in Eq. (10), and we focus on the values from *The Ivy* and *The Oxo Tower* restaurants. The criterion that has the greatest positive impact for *The Ivy* is *food*, so it is the *BestCriterion*. The CE matrix (see Fig. 4) shows the collective rating for that criterion is 8.9 in the interval $[0, 10]$. Regarding *The Oxo Tower*, we note that *drinks* criterion is the most detrimental, so it is the *WorstCriterion*.
- **Stage 2: Identifying relevant aspect terms through subgroup discovery.** We apply the Apriori-SD algorithm to the $BOC_4(food)$ table associated to *The Ivy* and extract the association rules with *positive* consequent in order to identify the most beneficial aspect terms referring its best criterion, the *food*. We also apply Apriori-SD to the $BOC_1(c_j)$, $j = 1, \dots, m$ tables associated to *The Oxo Tower* and extract the rules with *negative* consequent to identify all the aspect terms that are negatively evaluated by the experts. Table 3 lists the rules for both types of extracted rules. For example, the rule *pie* → *positive* indicates that pie is one of the most positively valued foods for *The Ivy*. The rules of each kind of subgroup, the positive and negative ones, are sorted in descending order based on the NWRAcc metric since it is the one considered at the Apriori-SD rule subset selection. We collect relevant aspect terms avoiding trivial ones such as *meal* and *menu* for the *food* criterion. Then, the *PositiveAspects* set contains *pie*, *crab*, *steak tartare*, and *liver*. *The Oxo Tower* has only four rules with *negative* consequent, which again proves that it is also high quality. Its corresponding *NegativeAspects* set is composed of *acoustics*, *manager*, and *waiter* since *drinks* is dropped for being a trivial aspect term for the *drink* criterion.
- **Stage 3: Identifying relevant sentences through attention mechanisms.** This stage identifies the sentences that have the most positive influence for *The Ivy* and the most negative influence for *The Oxo Tower* using the attention mechanism from the ASAM model. It has two attention layers that provide an attention weight for each input word. For example, Fig. 6 shows the attention weights that it assigns to an input sentence from *The Oxo Tower* restaurant. The attention layer from *the aspect and category processing block* focuses on the nouns *service* and *desserts*, which actually matches with aspects terms of two different opinions. The attention layer from *the polarity processing block* highlights the adjectives *great* and *delicious*, which represent positive sentiments. We average both kinds of weights and compute the sentence weight as shown in Eq. (13) for all the opinion sentences provided by the experts.
  We generate two rankings of positive sentences for *The Ivy* in order to have varied judgments, one about the best criterion, the *food*, and another about the overall restaurant. They are shown in Table 4 as Ranking 1 and Ranking 2, respectively. The first sentences of both rankings constitute the *Ranking* shown in the positive explanatory text. Additionally, we generate a ranking with negative judgments for *The Oxo Tower* restaurant without specifying any criteria, which is shown as Ranking 3 of Table 4.

Fig. 7 consolidates the information generated through the three previous steps by showing the two explanatory texts automatically generated by the ECDM-SDAM methodology for the case study.

5. Conclusions

This paper proposes the ECDM-SDAM methodology as an a posteriori system to support people making decisions that automatically generates explanations to justify its achieved result and captures the

$$\begin{array}{l|cccccc|}
 & \text{restaurant} & \text{food} & \text{service} & \text{drinks} & \text{ambience} & \text{location} \\
\text{Oxo Tower} & 0.27 & 0.28 & 0.181 & 0.0283 & 0.075 & 0.02837 \\
\text{J. Sheekey} & 0.28 & 0.31 & 0.187 & 0.028 & 0.0776 & 0.028 \\
\text{The Wolseley} & 0.274 & 0.29 & 0.186 & 0.028 & 0.071 & 0.028 \\
\text{The Ivy} & 0.283 & 0.3 & 0.19 & 0.029 & 0.077 & 0.028
\end{array}$$

○ BestCriterion

○ WorstCriterion

Fig. 5. Weighted Collective Evaluation matrix with identified criteria from the explanations.

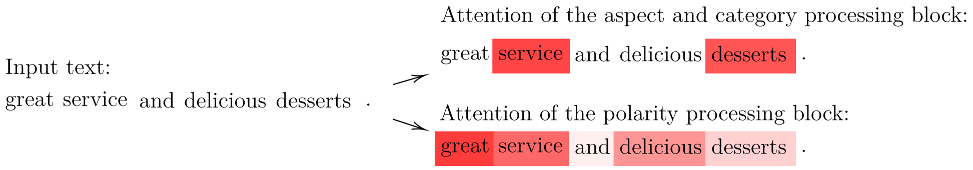
Attention of the aspect and category processing block:
great service and delicious desserts .
Input text:
great service and delicious desserts .
Attention of the polarity processing block:
great service and delicious desserts .

Fig. 6. Attention weights visualization of the two attention layers of the ASAM model. The intensity of the color indicates the relevancy or the attention value of the word.

Positive explanatory text

- The user should book at the restaurant *The Ivy* since it obtains the highest overall rating.
- Its criterion of greatest interest, *food*, reaches a rating of 8.9 out of 10.
- The pie, crab, steak tartare, and liver stand out positively.
- Two of the expert sentences that most benefit this restaurant being selected as the best are: *"good service and delicious food"* and *"constantly the best."*

Negative explanatory text

- The restaurant *The Oxo Tower* is high quality although it is the last one of the ranking, so we identify its weakest points.
- Its most detrimental criterion is *drinks*.
- The acoustics, manager, and waiter stand out negatively.
- Two of the expert sentences that most harm to this restaurant being selected as the last one are: *"poor service, meagre portions"* and *"not worth the trip"*.

Fig. 7. Explanatory texts provided by the ECDM-SDAM methodology to justify its final ranking.

Table 3
Significant rules extracted from the BOC tables using the Apriori-SD algorithm.

Restaurant	Rule	NWRAcc	Support	Confidence
The Ivy	*{meal}* → *positive*	0.52	0.027	1
The Ivy	*{pie}* → *positive*	0.51	0.022	1
The Ivy	*{crab}* → *positive*	0.51	0.022	1
The Ivy	*{menu}* → *positive*	0.51	0.022	1
The Ivy	*{steak tartare}* → *positive*	0.508	0.16	1
The Ivy	*{liver}* → *positive*	0.508	0.16	1
The Oxo Tower	*{drinks}* → *negative*	1	0.036	1
The Oxo Tower	*{acoustics}* → *negative*	0.75	0.023	1
The Oxo Tower	*{manager}* → *negative*	0.58	0.018	1
The Oxo Tower	*{waiter}* → *negative*	0.58	0.018	1

Table 4
Significant top 3 positive and negative sentences.

Ranking 1 Positive sentences for The Ivy (food criterion)	1. Good service and delicious food. 2. Food and service as always terrific. 3. The food, service and sense of occasion was truly perfect.
Ranking 2 Positive sentences for The Ivy (restaurant criterion)	1. Constantly the best 2. Always very good 3. Will definitely revisit for a special occasion.
Ranking 3 Negative sentences for The Oxo Tower (any criterion)	1. Not worth the trip 2. Poor service, meagre portions 3. Overall, therefore, it is poor value and plays to the tourist market.

wisdom of crowds available on social media. We consider opinions from unconstrained texts posted on social media as expert evaluations and design the explainable decision procedure.

The methodology incorporates two key novel components. Firstly, the ASAM neural network extracts the expert opinions and allows to identify the meaningful sentences provided by the experts through its attention mechanism. Secondly, the BOC tables represent the opinions and allow to identify the most relevant aspect terms associated with each criterion and alternative by applying SD algorithms on them. The methodology also identifies the most outstanding criteria and offers a positive explanation for the best alternative in the ranking as well as a negative explanation for the worst.

The experimental study manifests that the ECDM-SDAM methodology is adequate for solving decision problems from a practical point of view, since it analyzes the real data from the TripR-2020Large dataset. It is competitive and adequate as its achieved ranking matches the ranking obtained in [2]. In that previous study only the final ranking is provided, whereas the ECDM-SDAM methodology is able to explain why that result is obtained which makes it much more reliable.

The experimental study may state the following conclusions:

1. The explainability of DM models is essential for their wider use.
2. The ECDM-SDAM methodology provides easily understandable explanations of its sense-making mechanism.
3. Attention mechanisms and SD techniques are suitable for designing explainable DM models.
4. The ECDM-SDAM methodology captures the wisdom of crowds and can handle quality unconstrained natural evaluations using SA.

5. The ASAM model is competitive since it results on the TripR-2020Large and SemEval-2016 datasets outperforms previous studies.

As future work we plan to explore different areas of work. First, regarding to the database, it is necessary to create more datasets for CDM with natural language opinions to encourage the development of new studies to take advantage of this profitable environment where natural representations of opinions are offered. Second, another interesting avenue would be to integrate large language models (LLM) to design hybrid evaluation models since they have achieved a remarkable performance on different NLP tasks. A possibility is to use nanoPALM, a model inspired by nanoGPT, that achieves state-of-the-art few-shot results across hundreds of tasks with a high efficient training of very large neural networks [40]. Others models that can be integrated are SpikeGPT [41] and LLaMA [42]. We will intend to use LLMs to summarize the evaluations and to provide relationships between them.

CRediT authorship contribution statement

Cristina Zuheros: Conceptualization, Methodology, Investigation, Resources, Software, Writing – review & editing. **Eugenio Martínez-Cámara:** Conceptualization, Methodology, Investigation, Writing – review & editing. **Enrique Herrera-Viedma:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Iyad A. Katib:** Conceptualization, Supervision. **Francisco Herrera:** Conceptualization, Methodology, Writing – review & editing, Supervision.

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.

Data availability

I have shared the link to my data at the manuscript.

Acknowledgments

This work was partly supported by the grants PID2020-119478GB-I00, PID2019-103880RB-I00 and PID2020-116118GA-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". C. Zuheros is supported by the grant PRE2018-083884 funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future". Funding for open access charge: Universidad de Granada / CBUA

References

[1] Bin Zhu, Zeshui Xu, Jiuping Xu, Deriving a ranking from hesitant fuzzy preference relations under group decision making, IEEE Trans. Cybern. 44 (8) (2013) 1328–1337.

[2] Cristina Zuheros, Eugenio Martínez-Cámara, Enrique Herrera-Viedma, Francisco Herrera, Crowd decision making: Sparse representation guided by sentiment analysis for leveraging the wisdom of the crowd, IEEE Trans. Syst. Man, Cybern. Syst. 53 (1) (2023) 369–379.

[3] Tong Zhang, Xinrong Gong, C.L. Philip Chen, BMT-Net: Broad multitask transformer network for sentiment analysis, IEEE Trans. Cybern. 52 (7) (2022) 6232–6243.

[4] Mengqi Li, Yejun Xu, Xia Liu, Francisco Chiclana, Francisco Herrera, A trust risk dynamic management mechanism based on third-party monitoring for the conflict-eliminating process of social network group decision making, IEEE Trans. Cybern. (2022) http://dx.doi.org/10.1109/TCYB.2022.3159866, in press.

[5] Peijia Ren, Zeshui Xu, Zhinan Hao, Hesitant fuzzy thermodynamic method for emergency decision making based on prospect theory, IEEE Trans. Cybern. 47 (9) (2016) 2531–2543.

[6] Zhichao Wang, Yan Ran, Chuanxi Jin, Yifan Chen, Genbao Zhang, An additive consistency and consensus approach for group decision making with probabilistic hesitant fuzzy linguistic preference relations and its application in failure criticality analysis, IEEE Trans. Cybern. 52 (11) (2022) 12501–12513.

[7] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al., Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.

[8] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, Dino Pedreschi, A survey of methods for explaining black box models, ACM Comput. Surv. 51 (5) (2018) 1–42.

[9] Julia Amann, Dennis Vetter, Stig Nikolaj Blomberg, Helle Collatz Christensen, Megan Coffee, Sara Gerke, Thomas K Gilbert, Thilo Hagendorff, Sune Holm, Michelle Livne, et al., To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems, PLoS Digit. Health 1 (2) (2022) e0000016.

[10] Auste Simkute, Ewa Luger, Bronwyn Jones, Michael Evans, Rhianne Jones, Explainability for experts: A design framework for making algorithms supporting expert decisions more explainable, J. Responsib. Technol. 7 (2021) 100017.

[11] James Surowiecki, The Wisdom of Crowds, Anchor, 2005.

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

[13] Neha Punetha, Goonjan Jain, Game theory and MCDM-based unsupervised sentiment analysis of restaurant reviews, Appl. Intell. (2023).

[14] Devendra Kumar Tayal, Sumit Kumar Yadav, Divya Arora, Personalized ranking of products using aspect-based sentiment analysis and Plithogenic sets, Multimedia Tools Appl. 82 (1) (2023) 1261–1287.

[15] Yucheng Zhu, Xuanhua Xu, Bin Pan, A method for the dynamic collaboration of the public and experts in large-scale group emergency decision-making: Using social media data to evaluate the decision-making quality, Comput. Ind. Eng. 176 (2023) 108943.

[16] Juan Antonio Morente-Molinera, Gang Kou, Yi Peng, Cristóbal Torres-Albero, Enrique Herrera-Viedma, Analysing discussions in social networks using group decision making methods and sentiment analysis, Inform. Sci. 447 (2018) 157–168.

[17] Juan Antonio Morente-Molinera, Gang Kou, Konstantin Samuylov, Raquel Ureña, Enrique Herrera-Viedma, Carrying out consensual group decision making processes under social networks using sentiment analysis over comparative expressions, Knowl.-Based Syst. 165 (2019) 335–345.

[18] Yuanyuan Liang, Yanbing Ju, Peiwu Dong, Xiao-Jun Zeng, Luis Martínez, Jinhua Dong, Aihua Wang, A sentiment analysis-based two-stage consensus model of large-scale group with core-periphery structure, Inform. Sci. 622 (2023) 808–841.

[19] José Ramón Trillo, Enrique Herrera-Viedma, Juan Antonio Morente-Molinera, Francisco Javier Cabrerizo, A large scale group decision making system based on sentiment analysis cluster, Inf. Fusion 91 (2023) 633–643.

[20] Boopathy Prabadevi, N Deepa, Kaliyaperumal Ganesan, Gautam Srivastava, A decision model for ranking Asian Higher Education Institutes using an NLP-based text analysis approach, ACM Trans. Asian Low-Resour. Lang. Inform. Process. 22 (3) (2023) 1–20.

[21] Enrique Herrera-Viedma, Iván Palomares, Cong-Cong Li, Francisco Javier Cabrerizo, Yucheng Dong, Francisco Chiclana, Francisco Herrera, Revisiting fuzzy and linguistic decision making: scenarios and challenges for making wiser decisions in a better way, IEEE Trans. Syst. Man, Cybern. Syst. 51 (1) (2020) 191–208.

[22] Albert Verasius Dian Sano, Adriel Anderson Stefanus, Elizabeth Paskahlia Gunawan, Proposing tourism chatbot by employing the wisdom of crowds in building its knowledge base, in: 2022 International Conference on Information Management and Technology, ICIMTech, IEEE, 2022, pp. 634–638.

[23] Jinyue Feng, Chantal Shaib, Frank Rudzicz, Explainable clinical decision support from text, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2020, pp. 1478–1489.

[24] Qiaoting Zhong, Xiuyi Fan, Xudong Luo, Francesca Toni, An explainable multi-attribute decision model based on argumentation, Expert Syst. Appl. 117 (2019) 42–61.

[25] Xiuyi Fan, Francesca Toni, Explainable decision making with lean and argumentative explanations, 2022, arXiv preprint arXiv:2201.06692.

[26] Yuzhu Wu, Zhen Zhang, Gang Kou, Hengjie Zhang, Xiangrui Chao, Cong-Cong Li, Yucheng Dong, Francisco Herrera, Distributed linguistic representations in decision making: Taxonomy, key elements and applications, and challenges in data science and explainable artificial intelligence, Inf. Fusion 65 (2021) 165–178.

[27] Dzmitry Bahdanau, Kyung Hyun Cho, Yoshua Bengio, Neural machine translation by jointly learning to align and translate, in: 3rd International Conference on Learning Representations, ICLR 2015, 2015.

[28] Andrea Galassi, Marco Lippi, Paolo Torroni, Attention in natural language processing, IEEE Trans. Neural Netw. Learn. Syst. 32 (10) (2020) 4291–4308.

[29] Wei Wu, Houfeng Wang, Tianyu Liu, Shuming Ma, Phrase-level self-attention networks for universal sentence encoding, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 3729–3738.

[30] Yequan Wang, Minlie Huang, Xiaoyan Zhu, Li Zhao, Attention-based LSTM for aspect-level sentiment classification, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 606–615.

[31] Rohan Kumar Yadav, Lei Jiao, Morten Goodwin, Ole-Christoffer Granmo, Positionless aspect based sentiment analysis using attention mechanism, Knowl.-Based Syst. 226 (2021) 107136.

[32] Ting Huang, Zhi-Hong Deng, Gehui Shen, Xi Chen, A window-based self-attention approach for sentence encoding, Neurocomputing 375 (2020) 25–31.

[33] Zhaoyang Niu, Guoqiang Zhong, Hui Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.

[34] Sofia Serrano, Noah A. Smith, Is attention interpretable? in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 2931–2951.

[35] Franciso Herrera, Cristóbal José Carmona, Pedro González, María José Del Jesus, An overview on subgroup discovery: foundations and applications, Knowl. Inf. Syst. 29 (3) (2011) 495–525.

[36] Miguel López, Eugenio Martínez-Cámara, M Victoria Luzón, Francisco Herrera, ADOPS: Aspect Discovery Opinion Summarisation Methodology based on deep learning and subgroup discovery for generating explainable opinion summaries, Knowl.-Based Syst. 231 (2021) 107455.

[37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[38] Branko Kavšek, Nada Lavrač, APRIORI-SD: Adapting association rule learning to subgroup discovery, Appl. Artif. Intell. 20 (7) (2006) 543–583.

[39] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, et al., SemEval-2016 task 5: Aspect based sentiment analysis, in: Proc. of the 10th International Workshop on Semantic Evaluation, SemEval-2016, Association for Computational Linguistics, 2016, pp. 19–30.

[40] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al., Palm: Scaling language modeling with pathways, 2022, arXiv preprint arXiv:2204.02311.

[41] Rui-Jie Zhu, Qihang Zhao, Jason K. Eshraghian, SpikeGPT: Generative pre-trained language model with spiking neural networks, 2023, arXiv preprint arXiv:2302.13939.

[42] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al., Llama: Open and efficient foundation language models, 2023, arXiv preprint arXiv:2302.13971.