Knowledge-Based Systems 231 (2021) 107455

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

ADOPS: Aspect Discovery OPinion Summarisation Methodology based on deep learning and subgroup discovery for generating explainable opinion summaries

Miguel López [a], Eugenio Martínez-Cámara [a,*], M. Victoria Luzón [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, Spain
[b] Department of Software Engineering, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, Spain

ARTICLE INFO

Article history:
Received 8 November 2020
Received in revised form 21 July 2021
Accepted 29 August 2021
Available online 1 September 2021

Keywords:
Opinion summarisation
Deep learning
Aspect extraction
Subgroup discovery
Interesting rules

ABSTRACT

Opinion summarisation is concerned with generating structured summaries of multiple opinions in order to provide insightful knowledge to end users. We present the Aspect Discovery for OPinion Summarisation (ADOPS) methodology, which is aimed at generating explainable and structured opinion summaries. ADOPS is built upon aspect-based sentiment analysis methods based on deep learning and Subgroup Discovery techniques. The resultant opinion summaries are presented as interesting rules, which summarise in explainable terms for humans the state of the opinion about the aspects of a specific entity. We annotate and release a new dataset of opinions about a single entity on the restaurant review domain for assessing the ADOPS methodology, and we call it ORCo. The results show that ADOPS is able to generate interesting rules with high values of support and confidence, which provide explainable and insightful knowledge about the state of the opinion of a certain entity.

1. Introduction

Sentiment Analysis is concerned with the computational treatment of opinion, sentiment and subjectivity in text [1,2]. The analysis of the subjective meaning may be performed at different granularity levels, for instance from document to aspect level [3,4], but it gives the view of only one single person, the opinion holder. Due to the subjective nature of opinions, the polarity value of one of them is insufficient to know the state of opinion regarding a certain entity. Hence, some form of summary or synthesis of the opinion is needed to aggregate the views of several people in order to understand the state of the opinion of a particular entity.

Opinion summarisation can be tackled from a linguistic point of view, and we may use text summarisation techniques to extract the most meaningful sentences (extractive summaries), or to generate a text that exclusively includes the main opinion meaning of the text (abstractive summaries) [5]. These methods build raw text summaries, which can be used to generate even more succinct summaries by using textual entailment techniques [6] or to generate text-based opinion knowledge graphs [7].

The opinion summaries generated by text-based opinion summarisation techniques are still raw text and lack sufficient structure. A structured opinion summary allows addition qualitative and quantitative analysis to be performed in order to give insights into which specific element of an opinion receives most of the negative opinions, or which is the element that people express their views the most. Hence, structured opinion summaries are needed to ease the understanding, interpretation and acquisition of novel knowledge from opinions for making decisions.

In this paper we claim that the generation of structured opinion summaries requires the combination of different computational techniques, and it has to be performed at aspect level. Accordingly, we first use aspect-based sentiment analysis methods (ABSA) [8] to extract the aspect terms mentioned in the opinions and consolidate them into aspect categories. Then, we use Subgroup Discovery (SD) techniques [9] to discover new knowledge and expose it in explainable terms for humans. We address the combination of ABSA and SD to generate structured opinion summaries using a new opinion summarisation methodology that we called **A**spect **D**iscovery for **OP**inion **S**ummarisation (ADOPS). Since the ADOPS methodology is based on ABSA, it allows us to know which aspects of a specific entity are receiving more opinions or, more insightfully for a product manufacturer, which aspects of a product have more negative opinions.

Opinion summaries have to ease the understanding of the state of the opinion about a particular entity. Likewise, those

* Corresponding author.
E-mail addresses: miguellopezcamp@ugr.es (M. López), emcamara@decsai.ugr.es (E. Martínez-Cámara), luzon@ugr.es (M.V. Luzón), herrera@decsai.ugr.es (F. Herrera).

https://doi.org/10.1016/j.knosys.2021.107455
0950-7051/

summaries have to be presented in interpretable terms for humans. Hence, some insights should be given about how they are elaborated, in the line of the new requirements of explainable artificial intelligence (XAI) [10]. Recent research on SA shows that there is increasing interest in improving the explainability of SA models and using SA datasets to evaluate explainable models. For instance, in [7] the authors propose the use of causal graphs to provide explainable opinion summaries. Other studies are focused on conducting some modifications to the models so as to improve their explainability, like [11] that proposes the Diversity Long-Short Term Memory Recurrent Neural Network, which generates more diverse hidden states that contribute to more explainable attention weights in SA problems. On the other hand, SA datasets are used in [12] to evaluate the simulability capacity of different explanation methods, such as LIME [13] and Anchor [14].

In regard to the interest in explainability in SA, the ADOPS methodology integrates SD techniques that are able to synthesise the state of the opinion with interesting rules. Those rules resemble human behaviour and they ease the understanding and interpretability of the opinion summaries to humans. Moreover, the resultant interesting rules are only composed by few aspect terms, for the sake of the clarity of the resultant summaries and the transparency of the methodology. Hence, the ADOPS methodology generates explainable summaries because:

1. It generates easy-to-understand rules in human terms that link aspect terms with opinion values.
2. The summaries are generated with algorithmic transparent methods, since the evaluation measures of SD techniques can be understood by humans.

The ADOPS methodology is independent of a specific implementation of an ABSA method or SD algorithm, thus we can use off-the-shelf solutions or develop ad hoc methods. In this paper, we implement the ADOPS methodology using the Attention-Based Aspect Extraction method (ABAE) [15], which is an unsupervised deep learning model used for aspect term extraction. Due to the unsupervised nature of ABAE, it returns too fine-grained aspects without taking into account the domain relationships among them and lacking of an automatic categorisation in aspect term categories. Hence, we use an adaptation of ABAE [16], which consists of initialising ABAE with some aspect terms used as seed words and we call it ABAE Aspect Seed ($ABAE_{AS}$). The initialisation with aspect seed words allows aspect terms that are related to the domain to be extracted and to be categorised in their respective aspect categories. Categorisation into aspect categories will allow SD techniques to discover more insightful new knowledge.

The aim of the SD step of the ADOPS methodology is the generation of the minimum and most informative rules for providing novel explainable knowledge. Accordingly, we evaluate the APriori-SD [17], which is a SD implementation of the APriori algorithm, and the NMEEF-SD algorithm [18], which is a multi-objective evolutionary algorithm for SD. The NMEEF-SD algorithm is able to optimise several quality measures for SD, which means that the algorithm has the capacity of providing fewer yet higher quality rules.

The ADOPS methodology, as a methodology for opinion summarisation, requires a dataset of opinions that are focused on only one entity. Due to the lack of datasets that clusters opinions of one entity, we elaborate a new one focused on a single entity, which we called OneRestaurant Corpus (ORCo). The reviews of ORCo are from Tripadvisor,[1] and are about the restaurant review domain. Three annotators manually annotated the aspect categories mentioned in the opinions and the opinion meaning, and they reached high values of inter-annotator agreement according to the Kripendorff's α [19] and *multi*-κ [20] coefficients. Hence, ORCo will allow opinion summarisation methods on real opinions from social media sites to be evaluated.

Therefore, the main contributions of this paper are:

1. ADOPS, a methodology for generating structured and explainable summaries of the state of the opinion about the aspects of one entity.
2. ORCo,[2] a new dataset of opinions about one entity in the restaurant domain. Since the aspect categories and the opinion values were manually annotated, ORCo can be used to assess the extraction and categorisation of aspects in addition to the generation of opinion summaries.

We evaluate the structured opinion summaries of the ADOPS methodology with ORCo, and with two datasets from the domain of monument opinions to compare our structured opinion summaries with the ones generated in [21]. Likewise, we analyse the capacity of the ADOPS methodology to consolidate the lexical diversity of the aspect terms of an entity in a reduced set of aspect categories and how the generated opinion summaries encompass the opinions of the datasets.

The remainder of this paper is organised as follows: Section 2 presents propaedeutic knowledge which helps to explain the ADOPS methodology, and some related studies regarding opinion summarisation. Section 3 describes the ADOPS methodology, and Section 4 presents the ORCo dataset. Section 5 presents the evaluation of the ADOPS methodology, and Section 6 discusses the qualitatively analyses of the results. Finally, Section 7 presents the conclusions and future work.

2. Background

The ADOPS methodology aims at summarising the views about one entity by using interesting rules. These rules will be generated upon the categories of the aspect terms that are mentioned in the opinions. Hence, it will require the combination of methods from aspect-based opinion summarisation and SD.

In Section 2.1, we introduce the task of opinion summarisation, the methods related to SD are explained in Section 2.2, and we describe related approaches to generate explainable and structured summaries in Section 2.3.

2.1. Aspect-based opinion summarisation

The main characteristic of an opinion is its subjective nature, which makes it informative in the sense that it gives the personal view of the author. However, it cannot be considered conclusive because it only reflects the view of one person. For this reason, the task of opinion summarisation attains to synthesise the view expressed in several opinions [8]. The summary of several opinions resembles the task of text summarisation, or more specifically multi-document text summarisation [5].

Opinion summarisation is different from text summarisation in the sense that the summaries of opinions have to be focused on synthesising the subjectivity given by different people about one entity. In [22], the authors propose the use of a generative text model for building opinion summaries in raw text. In contrast, the authors of [16,23] advocate the use of extractive summarisation methods for also generating raw text summaries. However, we claim that the output of an opinion summary needs to be

[1] https://www.tripadvisor.com/.

[2] ORCo is available at https://github.com/ari-dasci/One-Restaurant-Corpus.

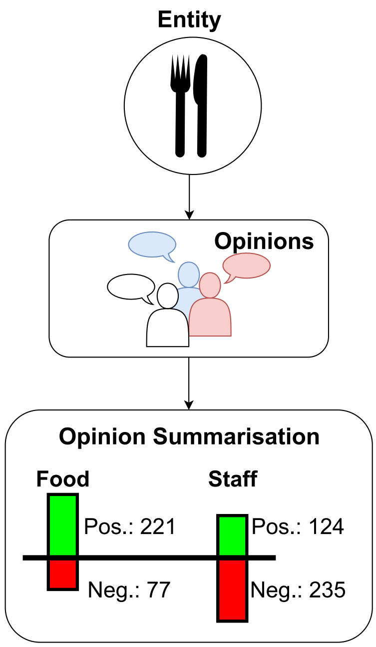

Fig. 1. This scheme represents the general workflow of an opinion summarisation methodology, which consists of extracting structured knowledge from a set of opinions about an entity.

structured, rather than being a raw text without any structure or with some predefined syntactic structure. Structured opinion summaries allow additional analyses that provide additional insights on the state of the opinion of one entity. ADOPS is able to build structured summaries by generating rules and measures that relate the mention of aspects to the state of the opinion.

Aspect-based opinion summarisation concerns with the generation of a structured opinion summary from a set of opinions about one entity. It has two main characteristics: (1) it is focused on the opinion targets, *i.e.* on the opinion about the entities or the aspects of those entities; and (2) it aggregates the opinions about the target with the aim of providing a qualitative opinion synthesis. Fig. 1 depicts the aspect-based summarisation task.

The extraction of the target opinion is built upon the ABSA task, which aims at identifying the opinion value about any entity and each of their aspects mentioned in the review [8]. ABSA is grounded in the definition of an opinion as the quintuple $(e_i, a_j, p_k, h_l, t_z)$, where e_i refers to the entity, a_j represents any aspect of the entity e_i, p_k the polarity value of the opinion about the aspect a_j of the entity e_i, h_l is the opinion holder or author of the opinion and t_z represents the time when the opinion was published. Therefore, the end goal of ABSA is the identification of at least the opinion value p_k of each aspect a_j of each entity e_i.

The ABSA task is usually split into three subtasks in order to make it more affordable. According to [24,25], the first one is the aspect term extraction (ATE) task, which is very similar to the information extraction task [26]. The ATE task identifies the spans of text that have an entity or an aspect meaning in the domain of the opinions. For instance, in the restaurant review domain the aim is to extract all the mentions (single or multi-word) of the target restaurant or its aspects. Once the aspect terms are identified, the next subtask is aspect category detection (ACD), roughly speaking, the categorisation of all the aspect terms into aspect categories. The subsequent tasks are related to the calculation of the opinion value of each aspect term (aspect term polarity) or of each aspect category (aspect category polarity). Regardless of the opinion holder and time, these three subtasks provide most of the opinion elements of ABSA, that is, the aspect terms, the categories of the aspect terms and the value of the opinions about them.

Concerning the definition of opinion summarisation, ADOPS is an aspect-based opinion summarisation methodology that can integrate the four subtasks of ABSA, as we describe in Section 3. In this paper, we implement the ADOPS methodology by only tackling the ATE and ACD subtasks, *i.e.* we extract the aspect terms and we categorised them into aspect categories of interest. Hence, we call aspect extraction (AE) phase to the ATE and ACD subtasks in the ADOPS methodology.

As mentioned above, the AE task resembles the information extraction task. AE can be addressed with a supervised or unsupervised approach. Following a supervised approach, there are models built upon the conditional random fields algorithm (CRF) [27], as in [28–31], and other studies based on the latest advances in deep learning with black box models [3,15,32,33] or deep learning methods with some explainable capabilities as [34].

Regarding unsupervised approaches, there are methods that are based on linguistic rules, such as [35], which is based on finding out frequent nouns, and [36], which leverages syntactic dependency information. In [37], the authors combine linguistic rules based methods and deep learning to obtain the aspect terms mentioned in opinions. On the other hand, aspect categories may be seen as topics because a review may be about, for instance, the food or the service of a restaurant. Accordingly, topic modelling techniques can be used to identify the different aspect terms and then to categorise them into topics or aspect categories. The first proposals were based on variants as extensions of the LDA algorithm [38], like [39–42]. The development of topic models based on deep learning, neural topic models [43,44], has pushed the development of neural models in which continuous representations of aspect categories are learnt. For instance, ABAE [15] is a neural topic model based on an autoencoder that builds continuous representations of aspect categories. We adapt ABAE following the same approach as [16] to implement the ABSA phase of the ADOPS methodology. The adaptation is based on the initialisation of the model with some aspect terms used as seed words with the aim of introducing some domain knowledge to the model, and we called this adaptation $ABAE_{AS}$. An additional variation of ABAE built upon an attention layer is described in [45]. Other adaptation of ABAE is presented in [23] where the authors leverage external knowledge in order to initialise the aspects embeddings.

2.2. Subgroup discovery

SD techniques are part of the supervised descriptive pattern mining field [46,47]. SD techniques extracts relationships between features that explain the behaviour of statistically interesting subgroups of individuals with respect to a target property of interest [9]. These techniques provide an informative set of interesting rules and metrics, which in conjunction give a structured description of the domain to be studied. These rules are composed of an antecedent or condition, which is a set of variables, and a consequent that is the target property of interest.

Since SD centres its analysis on a target property of interest, it is considered to be a field that combines predictive and descriptive paradigms. Thus, SD does not pretend to offer high predictive quality, but it gives an explanatory view of how variables affect to a target property in subgroups of individuals. Accordingly, SD techniques can be considered algorithmically transparent methods, because their quality measures are easy to understand for humans. Hence, SD methods have been also used for developing explainable models [48].

A rule generated by an SD algorithm can be interpreted as a set of co-appearances of features that define a statistically interesting subgroup of individuals. There is a wide range of SD algorithms, which may be classified as:

- **Classification based algorithms**, which are extensions of classical classification methods. For instance the Expert-guided SD algorithm [49] based on a variation of the beam search, or the CN2-SD [50], which is an adaptation of the classical CN2, a standard classification rule learning algorithm.
- **Association rule-based algorithms**, because of the similarities among association rule mining and SD. For instance the SD-Map method [51], which is an exhaustive search method that uses a minimum support threshold to reduce the search space; or the APriori-SD method [52] based on the APriori association rules mining algorithm. We use APriori-SD to implement the ADOPS methodology.
- **Evolutionary algorithms**, grounded in evolutionary methods that intend to get subgroups that optimise a set of quality metrics. MESDIF [53] and NMEEF-SD [18] are two examples of these kinds of SD algorithms. NMEEF-SD has been considered in this study.
- **Clustering approaches**, which are algorithms that adapt clustering techniques in order to address the SD task. For instance, [54] proposes an algorithm that extracts subgroups by using what they name cluster-grouping as a subtask in SD. In [55] it is proposed a clustering algorithm based on *k-medoids* specifically designed to address SD problems with many input and output features.

The evaluation measures of SD techniques assess the predictive capacity and the statistical interest of the generated interesting rules. These interesting rules has the form of $R : Cond \rightarrow Target_{value}$, where $Cond$ is the antecedent and $Target_{value}$ the consequent. Likewise, the evaluation measures take into account the frequency $n(x)$ of x in the dataset, the number of instances N of the dataset and the number n_c of possible values of $Target$. We subsequently describe the metrics used in this paper.

- **Support.** It measures the frequency of the co-occurrence of antecedents and consequents. Since SD is a descriptive mining technique, it can also be defined as the frequency of correctly classified examples according to the rule. It can be computed as:

$$\mathbf{Support}(\mathbf{R}) = \frac{n(Target_{value} \cdot Cond)}{N}$$

- **Confidence.** It is defined as the probability of getting the consequent if the antecedent is true:

$$\mathbf{Conf}(\mathbf{R}) = \frac{n(Target_{value} \cdot Cond)}{n(Cond)}$$

- **Normalised Weighted Relative Accuracy**. It is a modified version of Weighted Relative Accuracy (WRacc), also known as unusualness, which is defined as the balance between the rule's coverage:

$$Cov(R) = \frac{n(Cond)}{N}$$

and its accuracy gain:

$$AccG(R) = \frac{n(Target_{value} \cdot Cond)}{n(Cond)} - \frac{n(Target_{value})}{N}$$

$$WRAcc = Cov(R) \cdot AccG(R)$$

Its normalised version avoids WRAcc being conditioned to the percentage of the target variable and it is defined as:

$$\mathbf{NWRAcc} = \frac{WRAcc(R) - LB_{WRAcc}}{UB_{WRAcc} - LB_{WRAcc}}$$

being

$$UB_{WRAcc} = \frac{n(Target_{value})}{N} \cdot \left(1 - \frac{n(Target_{value})}{N}\right)$$

and

$$LB_{WRAcc} = \left(1 - \frac{n(Target_{value})}{N}\right) \cdot \left(0 - \frac{n(Target_{value})}{N}\right)$$

A *NWRAcc* greater than 0.5 is considered to have a good level of unusualness.

- **Significance.** It measures the significance of a finding as the likelihood ratio of a rule:

$$\mathbf{Sig}(\mathbf{R}) = 2 \cdot \sum_{k=1}^{n_c} n(Target_{valuek} \cdot Cond) \cdot \log \frac{n(Target_{valuek} \cdot Cond)}{n(Target_{valuek}) \cdot \frac{n(Cond)}{N}}$$

2.3. Structured summaries of opinions

We indicated in Section 2.1 that the end-goal of opinion summarisation is the generation of structured opinion summaries. However, most of the studies regarding opinion summarisation neglect the end goal, as they usually only perform a text summarisation. For instance, there are some studies that carry out extractive summaries that only return the most salient sentences with opinion information [16,56]. On the other hand, there are other studies that elaborate an abstractive summary from multiple opinions [57,58].

The development of a structured opinion summary requires the combination of several computational techniques from different tasks. Taking into account the strategy of aggregating the opinions after the extraction of the aspects, we have identified the following:

- Visualisation methods. After the extraction of the aspects, they visualise some statistics of the opinions about the different aspect terms. In [59], the authors utilise a variant of the LDA algorithm to extract the aspect terms, and then they show some statistic such as the main aspect categories, the sentences with the most salient aspect terms and their polarity. In [60], the authors combine a sequence to sequence deep learning model that performs an abstractive summary by depicting all the opinions as graph of aspect categories with the value of the opinion. The visualisation also includes the reasons behind the opinions gathered by the treatment of causal semantic relations [61], for instance "noisy rooms" are due to "heavy street noise".
- Knowledge graphs. The summary of a set of opinions can also be performed using the generation of a knowledge graph [62], which consolidates the semantic information of a set of documents, in this case opinions. In [7], the authors combine ABSA methods, identification of semantic causal relations and textual entailment to build a knowledge graph of opinions.
- Rule based methods. They describe relations between variables in an explainable way to humans, or more specifically, in this case they describe the relationship between an aspect category and the opinion value. Rule based methods have been used to make structured summaries since the proposal of the task in [35]. As we define in Section 2.2, SD methods also generate rules, but in this case the target is to know the antecedents stemmed from a specific consequent. In other words, given the negative opinion value as consequence, the aim is to find out which aspect categories, or antecedents, usually receive a negative opinion. In [21], the authors leverage SD methods to find out what aspect categories usually receive most of the negative opinions in the monument review domain.

The ADOPS methodology uses SD methods to structure the opinion summaries using interesting rules and it provides explainable opinion summaries. In contrast to similar approaches such as [21], the identification of aspect terms and their consolidation in aspect categories is performed by an unsupervised deep learning method that categorises the aspect terms into a reduced number of aspect categories, which allows to generate more insightful rules.

3. The ADOPS methodology

Obtaining explainable knowledge with regard to the comments made by both clients and users of a product is essential in terms of the decisions made by both clients and product manufacturers. Opinion summarisation concerns with obtaining a structured summary that extracts knowledge from a set of opinions with regard to an entity, centering the summary on explaining what users have expressed about the different aspects of the entity.

In this section we present the ADOPS methodology, which is aimed at building a structured opinion summary with respect to a single entity. We achieve this structured form by representing the summary as a set of interesting rules along with their quality measures obtained by SD techniques. Hence, these rules are formed by antecedents, which refer to the mention or 'not mention' of certain aspects in the opinions, and a consequent, which refers to the property of interest. In our study we use the polarity of the sentiment expressed in the opinions as the property of interest. Thus, opinion summaries built by the ADOPS methodology provide an explanation that links the mention of certain aspects to the sentiment associated with the opinions where those aspects appear.

Formally, the ADOPS methodology starts from a set of opinions D from different opinion holders with respect to a single entity e, where each opinion $d_i \in D$ is associated to a sentiment polarity $p_i \in \{-1, 1\}$ and a set of mentioned features or aspects a_{i_k} of e. Then, D can be represented in a structured way by means of considering each d_i as a set of mentioned a_{i_k} and its polarity p_i, naming this structured representation $D_{itemset}$. Therefore, from $D_{itemset}$, the ADOPS methodology aims at building an opinion summary using SD techniques, which represent the summary as a set of interesting rules. Broadly speaking, the ADOPS methodology takes a set of reviews with regard to a unique entity e as input, which can be considered to be unstructured data, and returns a structured summary by employing SD techniques.

The ADOPS methodology addresses the opinion summarisation task by combining two artificial intelligence tasks. One is the AE task from ABSA, which extracts the aspects terms mentioned in text. The other is the SD task, which aims to obtain a set of rules with regard to the polarity of the sentiment expressed and summarises the opinions in an explainable way. These two tasks can be tackled independently in different ways, which makes our approach modular.

In order to connect both tasks in a proper workflow, we need to add an intermediate step with the goal of adapting the output of AE task to the input of SD. Hence, the ADOPS methodology consists of the three following basic steps, which are also depicted in Fig. 2:

- Extraction of the aspects mentioned in each sentence (see Section 3.1).
- Grouping sentences and aspects by opinion, with the intention of representing the set of opinions as a set of items (see Section 3.2).
- Extraction of rules of interest using SD techniques from the itemset (see Section 3.3).

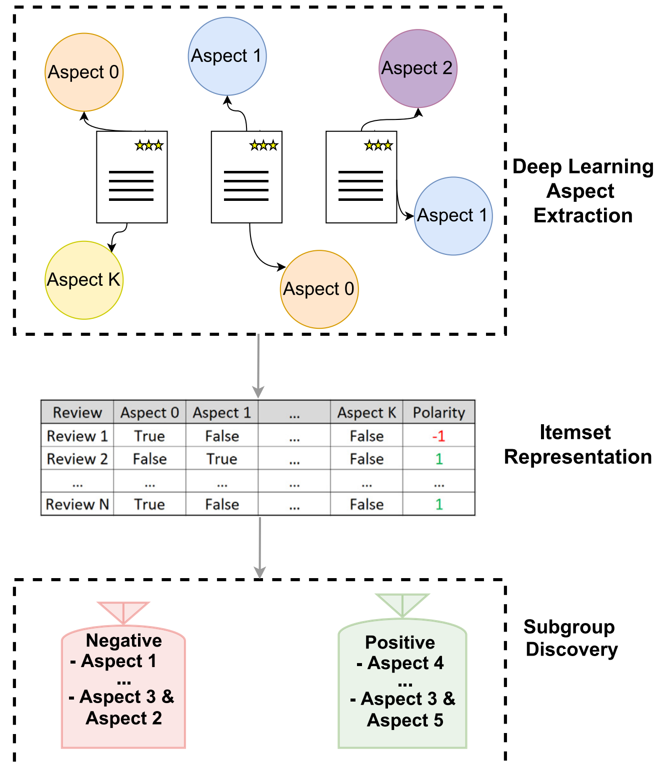

Fig. 2. It represents the ADOPS methodology. We start by extracting the aspects mentioned in all the opinions with deep learning techniques and we build an itemset from the extracted aspects. Finally, SD techniques relate the presence of certain aspects to the polarity of the opinions. This modular architecture allows to use different algorithms for the development of the methodology.

3.1. Aspect extraction

AE is a crucial step in aspect-based opinion summarisation, because it first identifies the aspect terms, and then consolidates them in aspect categories.

As we mentioned above, the ADOPS methodology is modular and may be implemented with different off-the-shelf or ad-hoc developed methods. The implementation of the AE step should match the following conditions:

1. To be an unsupervised approach. Accordingly, the model will be independent from the domain of the opinions, because it will not depend on the availability of annotated datasets.
2. To categorise the aspect terms into a reduced and informative set of aspect categories, which can be then used to perform additional analysis in conjunction with the opinion meaning.

In this paper, we implement the AE step with $ABAE_{AS}$, an adaptation of the ABAE method for aspect term extraction and aspect term categorisation. We describe the main characteristics of ABAE as follows:

ABAE. It is an unsupervised neural model approach which aims at finding the main topics or aspects mentioned in a set of texts [15]. Since it is an unsupervised algorithm, the only input required is a large non annotated set of text. In addition, this model can be easily guided by initialising the aspects matrix according to a set of seed words.

This model works as an autoencoder. First, ABAE codes a sentence as a weighted average word embedding, where the weights

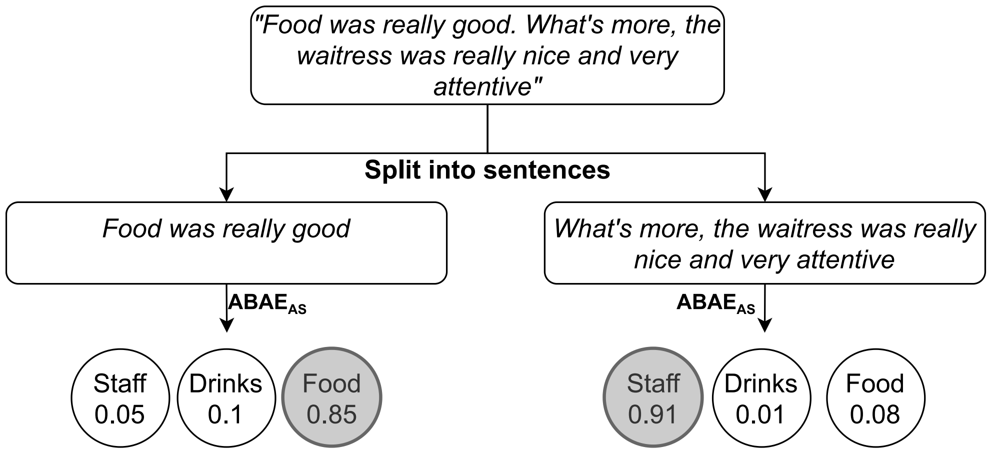

Fig. 3. Illustrative example of AE in our methodology applied to a specific opinion. Each sentence in an opinion is assigned an aspect according to the highest probability.

are provided by an attention layer. This attention layer gives more importance to the words that most influence the meaning of the sentence. Subsequently, the sentence is encoded as a probability distribution that gives the sentence the probability of belonging to each of the aspects categories. Then, the sentence is reconstructed by means of a product between the probability distribution and a matrix with K columns that represent each of the aspect categories of the domain of the opinions. The number of aspect categories K has to be defined.

ABAE offers two outputs of interest: (1) a vector continuous representation for each aspect category (aspects matrix), and (2) a probability distribution p_t over these K aspect categories, which classifies a sentence as one of the aspect categories. The aspects matrix that represents the aspect categories are originally initialised with the K-means algorithm.

Due to the unsupervised nature of ABAE, it frequently returns aspect terms that are too fine-grained. Accordingly, we use $ABAE_{AS}$, an adaptation proposed in [16], which consists in initialising the aspects matrix using a set of seed words per aspect. This initialisation guides the model in order to find more meaningful and homogeneous aspect categories. For instance, to initialise the *Food* aspect category, a possible set of seed words is *{food, cooked, meal, delicious, pizza}*.

We conduct the AE phase of the ADOPS methodology at sentence level, and we thus split each opinion into sentences. Hence, $ABAE_{AS}$ returns the p_t probability distribution over the aspect mentions of each sentence. We only consider one aspect per sentence, taking the highest probability as the correct aspect.

Fig. 3 shows an example of how AE works in the ADOPS methodology. The input opinion is split into sentences. Considering that the aspect matrix has been initialised with the aspects *Staff, Drinks* and *Food*, $ABAE_{AS}$ returns a probability distribution over them for each sentence. In this particular example, the sentence *Food was really good* gets a $p_{staff} = 0.05$, a $p_{drinks} = 0.1$ and a $p_{food} = 0.85$, indicating that the *Food* aspect is mentioned due to its higher probability.

3.2. Itemset representation

Once the aspect categories have been extracted from each sentence, we build the input of the SD method. To do so, we consider the aspect categories extracted as items that appear or do not appear in an opinion. Hence, our aim is to represent the set of opinions as an *itemset*.

For this purpose, we group the sentences and the aspects mentioned in these sentences by opinions. Then, we represent an opinion d_i as an array of binary discrete features $(a_{i_1}, a_{i_2}, \ldots, a_{i_k})$ and a polarity p_i, where $a_{i_j} \in \{0, 1\}, \forall j \in \{1, 2, \ldots, k\}$ depending on whether the presence or absence of the aspect a_j in the opinion d_i, and $p_i \in \{-1, 1\}$ corresponds to the polarity of the sentiment expressed in the opinion. Table 1 shows various opinions represented in an itemset form, where the Opinion Polarity column refers to the overall polarity of the opinion obtained by means of automatic or manual methods. The rest of the columns refer to the aspects to be analysed.

3.3. Subgroup discovery for structuring opinion summaries

The main goal of the ADOPS methodology is to summarise a set of opinions with regard to an entity in a structured way. We handle this by considering that there are relationships between aspects mentions that explain the sentiment polarity of an opinion.

These particularities make SD algorithms appropriate for this analysis. These techniques extract interesting rules, which explain the behaviour of a target property by means of patterns and relationships between variables of a subgroup of individuals. These algorithms give a set of rules which, along with their quality measures, give valuable descriptive and explainable information.

In order to extract interesting rules from the itemset generated, we consider the sentiment polarity of the opinion as the property of interest and the aspects mentioned in the opinion as the objects whose relationships explain the sentiment expressed. As result, we obtain a structured summary consisting on a set on rules and quality measures that summarise the sentiment with regard to a specific entity.

The structured summaries of ADOPS methodology are composed of rules. The antecedents of these rules refer to the aspects mentioned or not mentioned in the set of opinions, and the consequent is the sentiment associated to the opinions where these aspects are mentioned. Fig. 4 shows a specific example of rules that cover the two first examples from Table 1. These rules mean that when *Food* and *Staff* aspects are jointly mentioned, then the opinion is positive and that when *Drinks* is mentioned, then the opinion is more likely to be negative. In addition, these rules are accompanied by measures which quantify the covering of the rules. For example, the rule $\{Food = 1, Staff = 1\} \rightarrow \{Positive\}$ has a confidence of 0.73. This confident value means that 73 percent of the times that both aspects categories are mentioned, the associated sentiment is positive. Hence, *Food* and *Staff* usually receive positive opinions from the restaurant users.

Our SD problem is composed of binary discrete features that represent the presence or absence of an aspect in an opinion and

Table 1
Example of itemset representation of five opinions.

Opinion	Staff	Drinks	Food	Opinion Polarity
Food was really good. What is more, the waitress was really nice and very attentive.	1	0	1	1
Wine tasted like dirty water. I have never drunk anything worse.	0	1	0	−1
Wine list was very limited. Also, waitress treated us like we were children.	1	1	0	−1
You must try the steak.	0	0	1	1
The soup was disgusting. I told the waitress to take it away and she looked at me like I was a criminal. In addition, wine was not very tasty.	1	1	1	−1

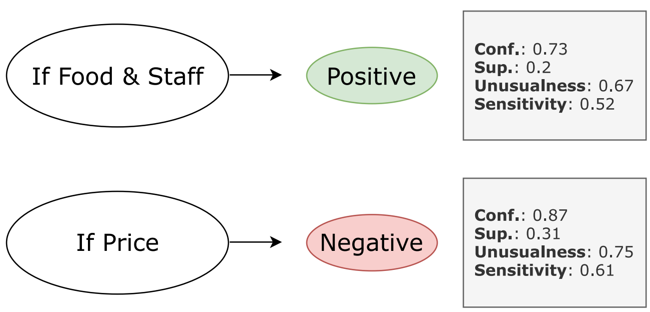

Fig. 4. Example of two rules and their respective measures that cover the two first opinions of the itemset shown in Table 1.

a binary target property, which refers to the polarity value of an opinion. To address this problem, due to the modular nature of the ADOPS methodology, we could use any SD algorithm. In this study we have considered two SD algorithms due to their respective features described as follows:

- APriori-SD [17] is a SD method based on APriori [63], a classical association rules method. APriori-SD incorporates a post-processing mechanism which filters out rules whose consequent is not the property of interest. APriori-SD, apart from taking into account confidence and support thresholds, employs a weighting post-processing scheme and a modification of WRAcc in order to get high quality rules.

  APriori-SD is an algorithm whose default input must be categorical binary features. For this reason, in addition to being a classical and quality SD method, we have chosen APriori-SD as one of the SD methods to be used in our study.
- Non-dominated Multiobjective Evolutionary Algorithm for Extracting Fuzzy Rules in Subgroup Discovery (NMEEF-SD) is a multi-objective evolutionary algorithm for SD based on fuzzy systems. This evolutionary algorithm employs a multi-objective optimisation in order to optimise multiple SD quality measures. This algorithm accepts numerical features without the need of a previous discretisation, due to numerical attributes are processed with a fuzzy system. Thus, classical evolutionary algorithms elements (chromosomes, mutation, crossover,etc.) are adapted to tackle the SD task.

  This evolutionary method is specifically focused on continuous features that are converted into linguistic fuzzy labels. Despite the fact that our opinions representation is built upon binary categorical features, we have used NMEEF-SD in our study because: (1) it extracts a smaller but more precise set of rules than other approaches, which eases the understanding of the resultant rules; (2) it also accepts categorical features as input; (3) it is a multi-objective approach, thus optimising various SD measures; and (4) its non-deterministic nature allows lower confidence and support thresholds to be explored with less computation time than deterministic models like APriori-SD.

4. ORCo: OneRestaurant corpus

Opinion summaries are focused on one single entity. Consequently, the evaluation of opinion summaries requires datasets where the opinions are clustered per entity. As far as we know, the available annotated datasets do not include the opinions clustered per entity, hence we would need this kind of resource in order to be able to evaluate the whole workflow of the ADOPS methodology.

We present ORCo, a new set of opinions referring to a single entity in the restaurant domain obtained from TripAdvisor. This dataset allows us to evaluate the whole workflow of the ADOPS methodology (the AE and SD steps). It contains 50 opinions divided in 277 sentences, with 25 opinions rated with 1 star and 25 with 5 stars in TripAdvisor. For each sentence, three annotators annotated the aspects which appear (Food, Drinks, Desserts, General, Staff, Ambience, Location and None) and the sentiment meaning (−1,0,1).

We evaluate the inter-annotator agreement for the annotation of the aspect categories and the sentiment meaning. We used Krippendorff's α [19] to assess the annotation of the aspect categories, reaching an α value of 0.7311, which is considered to be acceptable agreement because $\alpha \geq 0.667$. We also computed the *multi-*κ annotator coefficient [20] to evaluate the sentiment annotation, reaching a *multi-*κ value of 0.9041, considered to be an almost perfect agreement according to [64].

5. Experimental results

In this section, we evaluate the effectiveness of the ADOPS methodology.[3] First, we briefly describe the datasets used in our study, and present a classification of the datasets according to the requirements for evaluating each step of the ADOPS methodology in Section 5.1. Then, we show the $ABAE_{AS}$ results in the AE step of the ADOPS methodology as compared to two other similar approaches in Section 5.2. Finally, we analyse the rules that summarise the opinion sets obtained by two different SD algorithms in Section 5.3.

[3] We show the computation time of the steps of the ADOPS methodology in the Appendix.

Table 2
Domain train datasets employed to train the AE model and the embedding representations of words.

Datasets - Train	Sentences	Domain
TV Oposum Train	597,007	TV
Boots Oposum Train	423,754	Boots
Restaurant CitySearch Train	279,885	Restaurants
Monuments Train	234,246	Monuments

Table 4
SD evaluation datasets.

Datasets	Opinions	Pos. Rev.	Neg. Rev.	Sentences	Domain
SagradaFamilia	382	199	183	1723	Monuments
The Alhambra	239	120	119	1950	Monuments
ORCo	50	25	25	247	Restaurants

5.1. Datasets

We require datasets from different nature to evaluate each step of the ADOPS methodology. First, we need large non-annotated datasets of the target domains for generating domain word embeddings and for training $ABAE_{AS}$. Then, we require datasets with annotated aspect categories to evaluate the AE step of the ADOPS methodology. Finally, we use opinion sets about a single entity to properly evaluate the summaries generated by the SD step.

According to the aforementioned requirements we have employed the following opinion sets, which we have divided into three types of required sets:

- Boots and TV Oposum. Boots and TV opinions respectively proposed in [16]. They both consist of a non-annotated train set and an aspect annotated test set.
- Restaurant CitySearch. It is a restaurant opinions corpus [65] that consists of a non annotated train set and an aspect annotated test set.
- Sagrada Familia. TripAdvisor's opinions of the most popular tourist attraction in Barcelona. A unique dataset with TripAdvisor's ratings. Dataset proposed in [21]. Aspects are not annotated.
- The Alhambra. TripAdvisor's opinions of the most visited tourist attraction in Spain also proposed in [21]. It is an unique dataset also with TripAdvisor's ratings. Aspects are not annotated.
- ORCo. Opinions of a single restaurant proposed in this study. A unique set with aspects and sentiment annotated. In our study we considered TripAdvisor's ratings to be sentiment polarity of the opinions.

From these aforementioned opinion sets, we classify them in three different types in order to fulfil the requirements of each step of the ADOPS methodology previously discussed:

(1) *Domain train datasets*. In order to train word embeddings and the $ABAE_{AS}$ model, we need large non annotated datasets of the domain of interest. Table 2 shows the domain train sets employed in our analysis as well as the number of sentences and the domain of each dataset. TV and Boots Oposum Train refer to train partitions of TV and Boots Oposum datasets. The Restaurant CitySearch Train is the training partition of the Restaurant CitySearch dataset. The Monuments train dataset is built by joining *The Alhambra* and *Sagrada Familia*. All these datasets were preprocessed in order to remove stop words.

(2) *AE evaluation test datasets*. These are test sets with annotated aspects in order to evaluate the AE step of the ADOPS methodology. Table 3 displays a summary of the test sets employed to evaluate the AE step after removing the sentences without any aspect terms. Table 3 also shows the number of sentences, number of aspect labels, the domain of interest and all the aspect labels of these datasets. TV and Boots Oposum Test are the test partitions of the TV and Boots Oposum datasets after removing the sentences without any aspect terms. The Restaurant CitySearch Test is the test partition of the Restaurant CitySearch dataset. ORCo is the dataset presented in this study after removing sentences with the *None* aspect. To evaluate the models over these datasets, all stop words were removed.

(3) *SD evaluation datasets*. We use datasets about a single entity in order to get significant summaries. Table 4 summarises the datasets employed to evaluate SD, and it shows the number of opinions, number of positive and negative opinions, number of sentences and the domain of each dataset. We have considered the TripAdvisor's 5 stars rating to be positive polarity and the TripAdvisor's 1 star rating to be negative polarity, and we have filtered out the rest of the ratings. Also, we reduced the unbalance of the datasets *The Alhambra* and *SagradaFamilia* by means undersampling the reviews annotated with 5 stars.

5.2. Results of aspect extraction

We employ the *AE evaluation test datasets* to evaluate the AE step. We evaluate the quality of $ABAE_{AS}$ with two standard evaluation measures of classification tasks, namely Accuracy and Macro-average F1. We compare $ABAE_{AS}$ with two approaches that can also be initialised by means of seed words.

Table 3
AE evaluation test datasets.

Datasets - Test	Sentences	Test labels	Domain	Aspect Labels
TV Oposum Test	349	8	TV	Image, Sound, Connectivity, Ease of use, Price, Apps Interface, Customer Service, Size/Look.
Boots Oposum Test	325	8	Boots	Colour, Comfort, Durability, Look, Materials, Price, Size, Weather resistance.
Restaurant CitySearch Test	3,315	6	Restaurants	Food, Staff, Anecdotes, Ambience, Price, Miscellaneous.
ORCo	247	8	Restaurants	Food, Desserts, Drinks, General, Ambience, Staff, Location, Price.

Table 5
Parameters set for $ABAE_{AS}$.

$ABAE_{AS}$ Parameters	Value
Epochs	15
Batch size	50
Batches per epoch	100
Negative samples	20

Table 6
Test results of AE task.

Test Sets	$ABAE_{AS}$		K-Means		GuidedLDA	
	Acc.	M. F1	Acc.	M. F1	Acc.	M. F1
Restaurant CitySearch Test	**0.5427**	**0.3676**	0.4449	0.2793	0.5143	0.3384
TV Oposum Test	**0.6762**	0.6160	0.4699	0.4623	0.6446	**0.6166**
Boots Oposum Test	**0.4953**	**0.4643**	0.4461	0.3947	0.4769	0.4591
ORCo	**0.5263**	**0.3649**	0.4008	0.3024	0.4331	0.3198

- **K-Means** [66] is a classic unsupervised algorithm used for clustering whose centroids can be easily guided by initialising them with the desired values. Each centroid represents an aspect embedding. The aspect assigned to a sentence is the one whose word embedding is the closest to the sentence's average word embedding.
- **Guided-LDA** [67] is a modified version of *Latent Dirichlet Allocation* topic model. It incorporates lexical priors in order to skew the model to get topics according to the fed seed words. We set iterations parameter to 200 in our study.

We use the *word2vec* algorithm [68] to train the word embedding vectors on the *Domain Train Datasets* with an embedding size of 200, window size of 10 and negative sample size of 5. These word embeddings are used in $ABAE_{AS}$ and K-Means. The vocabulary is reduced to the 9,000 most frequent words. For K-Means and $ABAE_{AS}$, the initial aspect embeddings are initialised with the average word embedding of the seed words that correspond to that aspect. Same seed words are employed for all the algorithms. Table 5 shows the parameters set of $ABAE_{AS}$ in our study.

Since some of the sentences in the *AE evaluation test datasets* are annotated with more than one aspect category, we consider a prediction to be correct if the aspect categories predicted are in the set of true labels. Table 6 shows the results of the evaluation over the *AE evaluation test datasets*, training the models over the respective *Domain train datasets*. $ABAE_{AS}$ generally outperforms the other two algorithms except in Macro-F1 of TV Oposum Test set, where Guided LDA gets a slightly better score. These results assure us that $ABAE_{AS}$ can adequately implement the AE step in ADOPS methodology.

5.3. Rules extraction with subgroup discovery

In this section we show the results of the main rules obtained by the SD algorithms and their respective metrics, which are briefly described in Section 2.2. To do this, we have used the *SD evaluation datasets* after extracting the aspects of interest from them with $ABAE_{AS}$.

The target aspect categories for each dataset are the following:

- ORCo: Staff, Location, Ambience, Food, General, Dessert, Price and Drinks.
- The Alhambra and Sagrada Familia: Architecture, Location, Staff, Ticketing, Price, Tourism Resources, General and Queues.

In this study we use the APriori-SD and the NMEEF-SD algorithms. Table 7 shows the parameters used to run the APriori-SD algorithm. These parameters have been set to keep a trade-off between computation time and quality of the rules.

Table 7
APriori-SD parameters.

APriori-SD Parameter	Value
Min. Confidence	0.3
Min. Support	0.2
Modified WRAcc Threshold	0.25

Table 8
NMEEF-SD parameters.

NMEEF-SD Parameter	Value
Number Evaluation	10,000
Population Size	100
Mutation Probability	0.05
Crossover Probability	0.6
Min. Confidence	0.1

Table 9
Total number of rules obtained by each method in the experiments and average measures.

	Avg Sup.	Avg Conf.	Avg NWRAcc	Avg Sign.	Num. Rules
NMEEF-SD	0.2778	0.7984	0.6872	9.806	21
APriori-SD	0.2588	0.6291	0.5832	2.5954	219

To run NMEEF-SD we employed the SDEFSR R package[4] and in Table 8 the algorithm parameters setting is shown. As in the previous algorithm, the method parameters have been set to keep a balance between computation times and the quality of the rules. On the other hand, when it comes to setting our multi-objective optimisation function, to the best of our knowledge there are no current studies on determining the most suitable quality measures for an SD algorithm. Consequently, we studied several combinations of quality measures, and we saw that the combined optimisation of the quality measures *Unusualness* and *Significance* reaches the highest performance.

In the experiments, APriori-SD extracts 74 rules in ORCo, 70 rules in The Alhambra and 75 rules in Sagrada Familia. In our case, APriori-SD obtains large sets of rules, where some of them are valuable like {Drinks = 0, Price = 1} → {Negative} in ORCo, with high confidence and NWRAcc. However, it also obtains low-quality rules like {Location = 0} → {Negative} with a confidence and NWRAcc of 0.42 in ORCo.

On the other hand, NMEEF-SD obtains 7, 8 and 6 rules in the ORCo dataset, The Alhambra and Sagrada Familia datasets, respectively. This low number of rules is due to the fact that NMEEF-SD tends to extract smaller but with better quality sets of rules. For example, NMEEF-SD obtains the rule {General = 1, Staff = 0} → {Positive} with a value of confidence of 1 and a NWRAcc 0.625.

Table 9 summarises the rules obtained with APriori-SD and NMEEF-SD by showing the average measures of the rules extracted. As we have already highlighted, APriori-SD extracts a bigger set of rules than NMEEF-SD. This means that APriori-SD obtains lower average measures due to the excess of low-quality rules in the set of rules. Hence, we must postprocess these rules in order to get the best ones according to the desired measures. In addition to the smaller set of rules generated by NMEEF-SD, they also show less antecedents, which makes them more interpretable and explainable. Consequently, we consider that the results obtained by NMEEF-SD are of higher quality than APRIORI-SD.

[4] https://CRAN.R-project.org/package=SDEFSR.

Table 10
Rules extracted from ORCo and rules' metrics. SD Alg. column shows which algorithms have obtained each rule.

Rules	SD Alg.	Sup.	Conf.	NWRAcc	Sign.
{Drinks = 0, Price = 1} → {Negative}	Both	0.30	0.94	0.78	6.38
{Drinks = 0, Staff = 1} → {Negative}	Both	0.38	0.76	0.76	3.08
{Drinks = 0, General = 1} → {Negative}	APriori-SD	0.30	0.83	0.74	3.79
{Price = 1} → {Negative}	Both	0.32	0.76	0.72	2.63
{Drinks = 0, General = 1, Staff = 1} → {Negative}	APriori-SD	0.24	0.92	0.72	4.76
{Price = 0} → {Positive}	Both	0.4	0.69	0.72	1.86
{Drinks = 1} → {Positive}	Both	0.26	0.81	0.68	2.92
{Price = 0 , Desserts = 1} → {Positive}	APriori-SD	0.2	0.83	0.66	2.53
{Price = 0 , Location = 0} → {Positive}	APriori	0.2	0.83	0.66	2.53

Table 11
Rules extracted from The Alhambra dataset and rules' metrics (Archit. refers to Architecture aspect). SD Alg. column shows which algorithms have obtained each rule.

Rules	SD Alg.	Sup.	Conf.	NWRAcc	Sign.
{Staff = 1} → {Negative}	Both	0.43	0.70	0.75	10.35
{Staff = 1, General = 0} → {Negative}	Both	0.34	0.76	0.73	13.09
{Staff = 1, Ticketing = 1} → {Negative}	Both	0.34	0.74	0.72	11.65
{Archit. = 0, Staff = 1} → {Negative}	Both	0.35	0.72	0.72	10.36
{Archit. = 0, Staff = 1, General = 0} → {Negative}	APriori-SD	0.29	0.79	0.71	10.2
{Staff = 0} → {Positive}	Both	0.31	0.82	0.75	17.82
{Staff = 0 & Price = 0} → {Positive}	APriori-SD	0.27	0.84	0.72	16.78
{Ticketing = 0 & Price = 0} → {Positive}	APriori-SD	0.20	0.74	0.63	6.58
{Ticketing = 0} → {Positive}	APriori-SD	0.25	0.68	0.63	4.79
{General = 1, Staff = 0} → {Positive}	NMEEF-SD	0.13	1.0	0.63	17.95

Table 12
Rules extracted from Sagrada Familia dataset and rules' metrics (Archit. and T. Res. refer to Architecture and Tourism Resources aspects). SD Alg. column shows which algorithms have obtained each rule.

Rules	SD Algorithm	Sup.	Conf.	NWRAcc	Sign.
{Staff = 1} → {Negative}	Both	0.27	0.68	0.66	11.44
{Staff = 1, General = 0} → {Negative}	Both	0.21	0.74	0.65	12.94
{Staff = 1, T.Res. = 0} → {Negative}	APriori-SD	0.21	0.70	0.63	9.91
{Staff = 1, Price = 0} → {Negative}	APriori-SD	0.20	0.70	0.63	9.87
{General = 0, Archit. = 0} → {Negative}	APriori-SD	0.20	0.62	0.59	4.37
{Staff = 0} → {Positive}	Both	0.40	0.65	0.66	7.14
{Staff = 0, Price = 0} → {Positive}	Both	0.31	0.68	0.65	7.91
{Staff = 0 , Archit. = 1} → {Positive}	APriori-SD	0.25	0.68	0.62	6.11
{Staff = 0, Price = 0, Archit. = 1} → {Positive}	APriori-SD	0.21	0.70	0.61	6.56
{Staff = 0, T.Res. = 0, Price = 0} → {Positive}	APriori-SD	0.24	0.65	0.59	3.78

Table 10 shows some interesting rules about ORCo like the first 3 rules, which indicate that the *Drinks* aspect is not mentioned in the opinions. These rules clearly show that the *Drinks* aspect is not related to negative opinions. This fact probably means that *Drinks* is a positive aspect of the restaurant. Also, *Price* looks like a decisive aspect which clearly indicates a bad opinion of the restaurant.

We present the most insightful rules from The Alhambra and Sagrada Familia datasets in Tables 11 and 12, respectively. We can see some rules of interest, *i.e.*, in both datasets the *Staff* aspect appears as a very commonly mentioned aspect in negative opinions. Also, the fact that the *Architecture* aspect does not appear in the negative rules shown clearly indicates that the *Architecture* aspect is more closely related to positive opinions. We observe that positive rules are biased to aspects that are not mentioned ({*General* = *0*}, {*Staff* = *0* , *Price* = *0*}). The rule from Table 11 {*General* = *1*, *Staff* = *0*} → {*Positive*} proves how NMEEF-SD can extract quality rules in low *Support* thresholds, avoiding the high run times that these search spaces produce in deterministic algorithms like APriori-SD.

The aforementioned quality measures generally show high values. For instance, in ORCo all the confidence values obtained were greater than 0.6. In addition, we obtained rules with high NWRAcc, with all the rules shown obtaining a value of NWRAcc $\geq$ 0.55. This means that, according to [69], we get *Contrast Sets* defined by [70] as item sets which differ meaningfully in their distributions across groups of individuals. These groups, in our case, are defined by the sentiment value.

6. Qualitative analysis of the ADOPS methodology

In this section we discuss the results expounded in Section 5. Since we represent each aspect category as a vector, in Section 6.1 we have analysed the results to see if similar aspect categories are close among them. Section 6.2 shows if the opinions are really reflected in the rules obtained by ADOPS methodology. Despite showing how the ADOPS methodology is able to structurally represent knowledge from opinions, we perform an error analysis in Section 6.3. Finally, in Section 6.4, we compare our approach to a similar one aiming at contrasting our results.

6.1. Categorisation of aspect categories

The results of the AE step in Table 6 show that there is a lack of accuracy which most likely affects the summaries obtained. In this Section we analyse the results to see if the aspects categories learned by $ABAE_{AS}$ make sense semantically.

We make a visual analysis of the aspect embeddings that $ABAE_{AS}$ returns. Figs. 5–7 respectively show *Boots*, *TV* and *Restaurant* domains represented by the word embeddings of their vocabularies after a Principal Component Analysis reduction of dimensionality. These word embeddings are those that are closest to the aspect category embeddings. We see how points are clustered by their aspect category, and how they are closer to other similar aspect clusters than semantically different aspect categories.

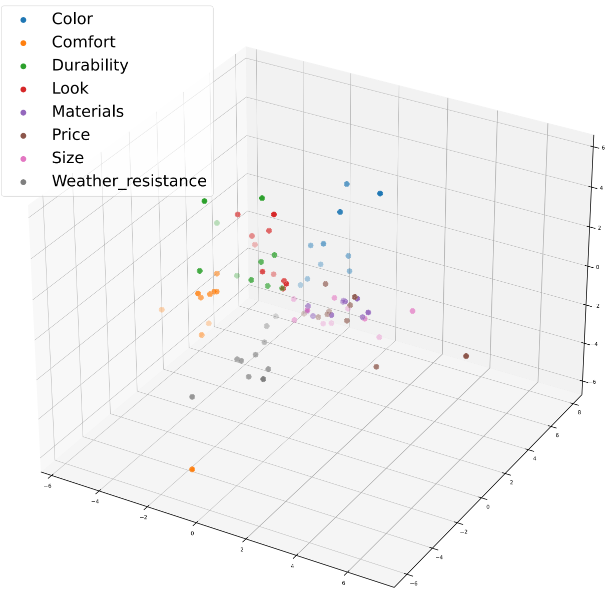

Fig. 5. 3D representation of the word embeddings closest to the aspects embeddings after a dimension reduction in the *boots* domain.

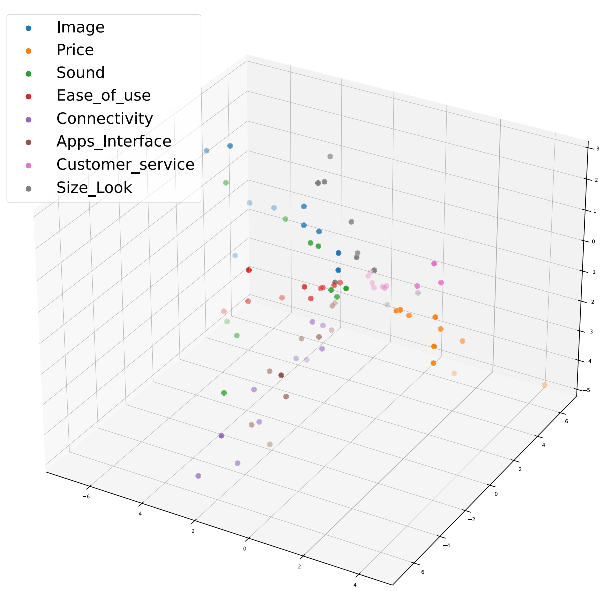

Fig. 6. 3D representation of the word embeddings closest to the aspects embeddings after a dimension reduction in the *tv* domain.

For instance, Fig. 5 shows grouped aspects in relation to aesthetic (*Colour*, *Look*). In addition, we see something similar in Fig. 6, where technical aspects like *Image* have other technical aspects nearby like *Sound*. Furthermore, in Fig. 7 we can see how the *Desserts* and *Food* clusters are closer together than other aspects like *Ambience*, which are far away. Also, the *General* aspect is in the middle, indicating that it does not specifically refer to anything.

Therefore, we can see how, despite the lack of accuracy of the AE step in the ADOPS methodology, the facts analysed from the graphics prove how the aspect embeddings adjusted by $ABAE_{AS}$ have semantic meaning.

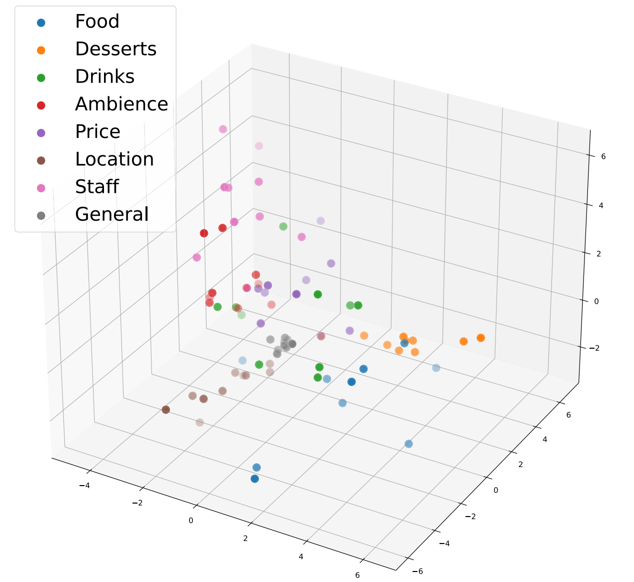

Fig. 7. 3D representation of the word embeddings closest to the aspects embeddings after a dimension reduction in the *restaurants* domain.

6.2. Analysis of how rules represent the opinions

The results previously given in the SD step show how the methodology is able to represent a structural and explainable summary of the opinions with regard to an entity. These rules are accompanied by measures, which provide a high interpretation capacity. However, in this Section we aim to determine whether the rules obtained really reflect the opinions about an entity.

Table 13 shows sentences of opinions covered by some of the rules obtained. For example, the rule $\{Staff = 1\} \rightarrow \{Negative\}$ of both *The Alhambra* and *Sagrada Familia* datasets covers clear sentences where the *Staff* aspect is mentioned and the sentiment is *Negative*. In addition, we show an example of the covering of a rule with more than one aspect used as antecedent, $\{Staff = 1, Ticketing = 1\} \rightarrow \{Negative\}$, in which we can see that an opinion is indeed covered when both aspects are clearly mentioned.

Rules of the ORCo dataset also present some sentences where the mention of the aspects are clear. For example, the rule $\{Drinks = 1\} \rightarrow \{Positive\}$, despite the *Drinks* aspect not being very common in the dataset, it covers opinions with correct mentions of the *Drinks* aspect, i.e., the sentence *"A wonderful selection of wines and some good advice"*.

Consequently, we confirm that the ADOPS methodology is able to generate, by means of rules and quality measures, an explainable and interpretable view of what users express about an entity.

6.3. ADOPS methodology error analysis

We show in Sections 5 and 6.2 that the ADOPS methodology generates valuable knowledge in the rules generated by SD. Nonetheless, we are aware that $ABAE_{AS}$ misclassified some aspect-categories, which implies that the SD method generates some rules that do not match the real meaning of the opinions.

As we have shown in Table 6, the results taken from the ORCo dataset are not high in the AE step, which means that there are some misclassified sentences. For example, the sentence *"dessert was really good though"*. is classified as *Drinks* probably due to the similarity among the aspect categories *Food, Drinks and Desserts*, whose contexts are probably similar in many cases.

Table 13
Examples of segments covered by rules.

ORCo	
"The gin and tonic was very small for the price we paid and presented very averagely." "Nobody would choose to pay these prices to sit in a box room upstairs." "Our meal was over £100 despite deciding to cut it short because of our upset."	{Price = 1} → {Negative}
"We had the 'Tu me manques' as a sharing cocktail before the tasting menu, all so good!" "A wonderful selection of wines and some good advice."	{Drinks = 1} → {Positive}
The Alhambra	
"I wish I had captured the name of the woman behind the ticket off but she took the biscuit for being so rude." "How is it possible that staff in one of Spain's largest tourist attractions get offended for being asked to speak English?" "The staff told us that it was our fault that we come so late - how brazen!"	{Staff = 1} → {Negative}
"Tickets were almost sold out so we had to use the 19:00 slot…The staff worked so slowly that we entered…"	{Staff = 1, Ticketing = 1}→ {Negative}
Sagrada Familia	
"The lady at the ticket office just told us that there will be no time for an audio tour but you can go around."	{Staff = 1} → {Negative}

In *The Alhambra* and *Sagrada Familia* datasets, we found some opinions which had been wrongly covered by rules. For instance, in the *Sagrada Familia* dataset, the opinion *"Wasted 1/2 h trying to throw at this purchasing tickets online. No luck for me. No bucks for them. Look at the outside only, move on to something functional".* appears as covered by the rule {Architecture = 1} → {Negative}. This error might be due to a certain bias of $ABAE_{AS}$ with respect to the prediction of the *Architecture* aspect, since the 23% of the sentences are classified into this aspect category.

Other errors are due to the presence of implicit aspects, which causes $ABAE_{AS}$ to fail. For instance, the sentence from ORCo dataset *"overall not coming back and only recommended if you want to see fake plants in a box room".* clearly rates the restaurant in general. However, $ABAE_{AS}$ classifies it as *Ambience*, probably biased by some explicit words such as *room* or *plants*, which are strongly related to the aspect category *Ambience*. This error causes the opinion where this sentence appears to be covered by the rule {Ambience = 1, Staff = 1} → {Negative}.

This lack of accuracy of the ADOPS methodology in the AE step indicates that there is still some room for improvement.

6.4. A comparative study

There are other approaches that also attain to build opinion summaries from a set of opinions. In this Section, we compare the ADOPS methodology with a similar approach presented in [21].

In [21], the authors perform the AE and SD steps with the aim of building rules that explain a set of opinions. They employ a supervised ATE model to address the AE task and the APriori-SD algorithm in the SD task. However, they do not successfully deal with the challenge of language diversity, *i.e.* the challenge of managing different mentions of the same meaning or concept, which in turn generates a sparse representation of aspect categories. They address this problem with the K-Means algorithm, but they do not completely resolve it. Table 14 shows the parameters used in [21].

Table 14
Parameters set-up used in [21].

AE Valdivia et al. Parameters	Value
Conv1 feature map	100
Conv1 filter size	2
Conv2 feature map	50
Conv2 filter size	2
Pool size	2
Apriori-SD Valdivia et al. Parameters	**Value**
Min. support	0.001
Min. confidence	0.01

They use *The Alhambra* and the *Sagrada Familia* datasets. The sparsity mentioned along with unbalanced data causes not quality rules. This causes the conclusions drawn to be not significant enough. Furthermore, their approach is supervised, making it dependent on annotated data.

Table 15 shows a comparison between some of the most outstanding rules obtained in [21] and some similar rules obtained by the ADOPS methodology. It shows the *Support, Confidence* and *WRAcc* of each rule, and the '-' symbol indicates that there are no measures available, because the respective approach did not generate that rule. We see how in *The Alhambra* dataset, we obtain rules that discuss similar aspects. They separately obtained *guard* and *staff* as common negative aspects. Since the ADOPS methodology categorises the aspect terms in fewer yet more informative aspect categories, the *Staff* aspect category encompasses the previous aspect categories. In the *Sagrada Familia* dataset, the extracted rules of the ADOPS methodology and [21] are quite different. However, the low quality values of the rules of [21] make it difficult to extract insightful conclusions.

We show how the ADOPS methodology extracts more generalisable rules. Besides this higher generalisation capacity, our proposal has the great advantage of not being dependent on an annotated dataset, causing it at the same time be less dependent on the domain of interest. Furthermore, the aspect categories can be defined by means the setting of aspect terms as seed words.

7. Conclusions

In this paper we propose using the ADOPS methodology to overcome the limitations of opinion summarisation in the generation of structured and explainable opinion summaries about

Table 15
Comparison of rules and measures extracted by NMEEF-SD in our study and in [21] using APriori-SD.

The Alhambra						
	ADOPS methodology			Valdivia et al. [21]		
Rule	*Sup*	*Conf*	*WRAcc*	*Sup*	*Conf*	*WRAcc*
{Staff = 1, Ticketing = 1} → {Negative}	0.34	0.74	0.11	–	–	–
{Queues = 1, Ticketing = 1} → {Negative}	0.24	0.58	0.04	–	–	–
{Staff = 1} → {Negative}	0.43	0.7	0.12	<0.01	0.28	<0.01
{Guard = 1} → {Negative}	–	–	–	<0.01	0.38	<0.01
{Queues = 1} → {Negative}	–	–	–	<0.01	0.15	<0.01
Sagrada Familia						
	ADOPS methodology			Valdivia et al. [21]		
Rule	*Sup*	*Conf*	*WRAcc*	*Sup*	*Conf*	*WRAcc*
{Staff = 1} → {Negative}	0.27	0.68	0.08	–	–	–
{Architecture = 0} → {Negative}	0.24	0.57	0.04	–	–	–
{Location = 0} → {Negative}	0.32	0.51	0.02	–	–	–
{Ceiling = 1} → {Negative}	–	–	–	<0.01	0.1	<0.01
{Natural = 1} → {Negative}	–	–	–	<0.01	0.07	<0.01
{Entry = 1} → {Negative}	–	–	–	<0.01	0.05	0

a single entity. The opinion summaries generated by the ADOPS methodology are built upon human interpretable rules. These rules link each aspect category with their most frequent opinion value, providing comprehensive knowledge to the target audience. In addition, the generation of summaries is conducted in an explainable way, in the sense that the interesting rules are built according to quality measures that are easy to understand for humans.

The ADOPS methodology is grounded in the combination of deep learning methods from ABSA for the extraction of aspect categories and SD techniques for generating the rules. Specifically, the ABSA step is implemented with the unsupervised deep learning method $ABAE_{AS}$, and the SD step with the NMEEF-SD algorithm, which returned fewer and more insightful rules than the APriori-SD algorithm, as shown in Table 9. The reduced set of rules returned by the NMEEF-SD algorithm contribute to the explainability of the ADOPS methodology, since it eases the understanding of the new knowledge to humans.

We highlight that the use of a deep learning model for the ABSA step permits obtaining a more adequate categorisation of aspects than other similar approaches, like [21] that is based on statistical number cluster setting methods. Accordingly, the ADOPS methodology can cluster the aspect terms in high-level aspect categories, reducing the lexical diversity of aspect terms and hence allowing SD techniques to discover more insightful knowledge.

The ADOPS methodology represents a first attempt to generate structured and explainable aspect based opinion summaries, with the aim of subsequently applying other data mining techniques and automatically generating reports about the state of the opinion of one entity. The ADOPS methodology can be combined with text based opinion summarisation methods in order to link the interpretable rules with a text opinion summary that consolidates all the opinions about an entity, as has been done in [7].

As future work, we will work on improving each element of the ADOPS methodology. We will work on the unsupervised extraction of not only explicit aspects but also implicit ones, as well as identifying all the aspect terms that are in a sentence. Likewise, we will work on enlarging ORCo with opinions from more entities, in order to compare the rules among entities, as well as adding opinions from a domain other than restaurant opinions in order to evaluate its multi-domain capacity.

CRediT authorship contribution statement

Miguel López: Conceptualization, Methodology, Investigation, Resources, Software, Writing – review & editing. **Eugenio Martínez-Cámara:** Conceptualization, Methodology, Investigation, Writing – review & editing. **M. Victoria Luzón:** Conceptualization, Methodology, Writing – review & editing, 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.

Acknowledgements

This research work is partially supported by the Trust-ReDaS (PID2020-119478GB-I00) and the FedDAP (PID2020-116118GA-I00) projects from the Spanish Government, the project DeepSCOP-Ayudas Fundación BBVA a Equipos de Investigación Científica en Big Data 2018 and a grant from the Fondo Europeo de Desarrollo Regional (FEDER). Eugenio Martínez-Cámara was supported by the fellowship programme Juan de la Cierva Incorporación (IJC2018-036092-I).

Table A.16
Computation time of the steps of the ADOPS methodology.

Step	Experiment	Time
Aspect extraction	$ABAE_{AS}$ Restaurant Domain	130.62 s
	$ABAE_{AS}$ Monuments Domain	204.56 s
	$ABAE_{AS}$ TV Domain	757.82 s
	$ABAE_{AS}$ Boots Domain	254.81 s
Extraction of interesting rules	Apriori-SD Sagrada Familia	8.28 s
	Apriori-SD The Alhambra	11.8 s
	Apriori-SD ORCo	37.8 s
	NMEEF-SD Sagrada Familia	58.03 s
	NMEEF-SD The Alhambra	49.87 s
	NMEEF-SD ORCo	37.8 s

Appendix. The computation time of the ADOPS methodology

The ADOPS methodology is based on 3 steps: (1) AE extraction step, (2) itemset representation of the set of opinions and (3) rules extraction through SD techniques. This workflow must be efficient enough in terms of time since it is a key factor to be considered for choosing a model for a specific task.

In order to give a notion of the efficiency of the ADOPS methodology, we show in Table A.16 the computation time in terms of seconds of the ADOPS methodology. The times shown correspond to the $ABAE_{AS}$ algorithm for aspect extraction and the two SD methods used here. All the experiments were run in an Intel Core i7-7700HQ CPU with 8 threads and 16 GB of RAM memory.

References

[1] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inf. Retr. 2 (1–2) (2008) 1–135.

[2] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107.

[3] S. Poria, E. Cambria, A. Gelbukh, Aspect extraction for opinion mining with a deep convolutional neural network, Knowl.-Based Syst. 108 (2016) 42–49.

[4] M.E. Basiri, S. Nemati, M. Abdar, E. Cambria, U.R. Acharya, ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis, Future Gener. Comput. Syst. 115 (2021) 279–294.

[5] A. Nenkova, K. McKeown, A survey of text summarization techniques, in: Mining Text Data, Springer US, 2012, pp. 43–76.

[6] I. Dagan, D. Roth, M. Sammons, F.M. Zanzotto, Recognizing textual entailment: Models and applications, Synth. Lect. Hum. Lang. Technol. 6 (4) (2013) 1–220.

[7] N. Carmeli, X. Wang, Y. Suhara, S. Angelidis, Y. Li, J. Li, W.-C. Tan, Constructing explainable opinion graphs from review, abs/2006.00119 (2021) arXiv:2006.00119.

[8] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, 2015.

[9] S. Ventura, J.M. Luna, Subgroup discovery, in: Supervised Descriptive Pattern Mining, Springer International Publishing, 2018, pp. 71–98.

[10] A.B. Arrieta, N. Díaz-Rodríguez, J.D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.

[11] A.K. Mohankumar, P. Nema, S. Narasimhan, M.M. Khapra, B.V. Srinivasan, B. Ravindran, Towards transparent and explainable attention models, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2020, pp. 4206–4216.

[12] P. Hase, M. Bansal, Evaluating explainable AI: Which algorithmic explanations help users predict model behavior? in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2020, pp. 5540–5552.

[13] M.T. Ribeiro, S. Singh, C. Guestrin, Why should I trust you?: Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, 2016, pp. 1135–1144.

[14] M.T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations, 2018.

[15] R. He, W.S. Lee, H.T. Ng, D. Dahlmeier, An unsupervised neural attention model for aspect extraction, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, vol. 1, 2017, pp. 388–397.

[16] S. Angelidis, M. Lapata, Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 3675–3686.

[17] B. Kavšek, N. Lavrač, V. Jovanoski, APRIORI-SD: Adapting association rule learning to subgroup discovery, in: Advances in Intelligent Data Analysis V, 2003, pp. 230–241.

[18] C.J. Carmona, P. González, M.J. Del Jesus, F. Herrera, Nmeef-SD: Non-dominated multiobjective evolutionary algorithm for extracting fuzzy rules in subgroup discovery, IEEE Trans. Fuzzy Syst. 18 (2010) 958–970.

[19] K. Krippendorff, Content Analysis: An Introduction To Its Methodology, Sage, 2004.

[20] M. Davies, J.L. Fleiss, Measuring agreement for multinomial data, Biometrics (1982) 1047–1051.

[21] A. Valdivia, E. Martínez-Cámara, I. Chaturvedi, M.V. Luzón, E. Cambria, Y. Ong, F. Herrera, What do people think about this monument? Understanding negative reviews via deep learning, clustering and descriptive rules, J. Ambient Intell. Humaniz. Comput. 11 (1) (2020) 39–52.

[22] A. Bražinskas, M. Lapata, I. Titov, Unsupervised opinion summarization as copycat-review generation, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 5151–5169.

[23] C. Zhao, S. Chaturvedi, Weakly-supervised opinion summarization by leveraging external information, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 9644–9651.

[24] J. Pavlopoulos, I. Androutsopoulos, Aspect term extraction for sentiment analysis: New datasets, new evaluation measures and an improved unsupervised method, in: Proceedings of the 5th Workshop on Language Analysis for Social Media, LASM, 2014, pp. 44–52.

[25] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, SemEval-2014 task 4: Aspect based sentiment analysis, in: Proceedings of the 8th International Workshop on Semantic Evaluation, SemEVal 2014, 2014, pp. 27–35.

[26] J.R. Hobbs, E. Riloff, Information Extraction, second ed., CRC Publisher, 2010, pp. 524–545.

[27] C. Sutton, A. McCallum, An introduction to conditional random fields, Found. Trends Mach. Learn. 4 (4) (2012) 267–373.

[28] B. Yang, C. Cardie, Joint inference for fine-grained opinion extraction, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2013, pp. 1640–1649.

[29] M. Chernyshevich, IHS R&D Belarus: Cross-domain extraction of product features using CRF, in: Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval 2014, 2014, pp. 309–313.

[30] Z. Toh, J. Su, NLANGP: Supervised machine learning system for aspect category classification and opinion target extraction, in: Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval 2015, 2015, pp. 496–501.

[31] D. Xenos, P. Theodorakakos, J. Pavlopoulos, P. Malakasiotis, I. Androutsopoulos, AUEB-ABSA at SemEval-2016 task 5: Ensembles of classifiers and embeddings for aspect based sentiment analysis, in; Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval-2016, 2016, pp. 312–317.

[32] W. Wang, S.J. Pan, D. Dahlmeier, X. Xiao, Coupled multi-layer attentions for co-extraction of aspect and opinion terms, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 3316–3322.

[33] X. Li, L. Bing, P. Li, W. Lam, Z. Yang, Aspect term extraction with history attention and selective transformation, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, 2018, pp. 4194–4200.

[34] Y. Ma, H. Peng, T. Khan, E. Cambria, A. Hussain, Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis, Cogn. Comput. 10 (4) (2018) 639–650.

[35] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2004, pp. 168–177.

[36] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, Comput. Linguist. 37 (1) (2011) 9–27.

[37] C. Wu, F. Wu, S. Wu, Z. Yuan, Y. Huang, A hybrid unsupervised method for aspect term and opinion target extraction, Knowl.-Based Syst. 148 (2018) 66–73.

[38] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (null) (2003) 993–1022.

[39] A. Mukherjee, B. Liu, Aspect extraction through semi-supervised modeling, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2012, pp. 339–348.

[40] Z. Chen, A. Mukherjee, B. Liu, Aspect extraction with automated prior knowledge learning, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014, pp. 347–358.

[41] S. Wang, Z. Chen, B. Liu, Mining aspect-specific opinion using a holistic lifelong topic model, in: Proceedings of the 25th International Conference on World Wide Web, 2016, pp. 167–176.

[42] M. Shams, A. Baraani-Dastjerdi, Enriched LDA (ELDA): Combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction, Expert Syst. Appl. 80 (2017) 136–146.

[43] Z. Cao, S. Li, Y. Liu, W. Li, H. Ji, A novel neural topic model and its supervised extension, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015, pp. 2210–2216.

[44] R. Ding, R. Nallapati, B. Xiang, Coherence-aware neural topic modeling, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 830–836.

[45] R. He, W.S. Lee, H.T. Ng, D. Dahlmeier, Effective attention modeling for aspect-level sentiment classification, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 1121–1131.

[46] W. Klösgen, Explora: A multipattern and multistrategy discovery assistant, in: Advances in Knowledge Discovery and Data Mining, American Association for Artificial Intelligence, 1996, pp. 249–271.

[47] S. Wrobel, An algorithm for multi-relational discovery of subgroups, in: European Symposium on Principles of Data Mining and Knowledge Discovery, 1997, pp. 78–87.

[48] C. Lonjarret, C. Robardet, M. Plantevit, R. Auburtin, M. Atzmueller, Why should I trust this item? Explaining the recommendations of any model, in: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA, 2020, pp. 526–535.

[49] D. Gamberger, N. Lavrac, Expert-guided subgroup discovery: Methodology and application, J. Artificial Intelligence Res. 17 (2011).
[50] N. Lavrač, B. Kavšek, P. Flach, L. Todorovski, Subgroup discovery with CN2-SD, J. Mach. Learn. Res. 5 (2004) 153–188.
[51] M. Atzmueller, F. Puppe, Sd-map – a fast algorithm for exhaustive subgroup discovery, in: Knowledge Discovery in Databases: PKDD 2006, 2006, pp. 6–17.
[52] B. Kavšek, N. Lavrač, APRIORI-SD: Adapting association rule learning to subgroup discovery, Appl. Artif. Intell. 20 (7) (2006) 543–583.
[53] M. del Jesus, P. Gonzalez, F. Herrera, Multiobjective genetic algorithm for extracting subgroup discovery fuzzy rules, in: 2007 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making, 2007, pp. 50–57.
[54] A. Zimmermann, L. De Raedt, Cluster-grouping: From subgroup discovery to clustering, Mach. Learn. 77 (2009) 125–159.
[55] L. Umek, B. Zupan, Subgroup discovery in data sets with multi-dimensional responses, Intell. Data Anal. 15 (2011) 533–549.
[56] V. Priya, K. Umamaheswari, Aspect-based summarisation using distributed clustering and single-objective optimisation, J. Inf. Sci. 46 (2) (2020) 176–190.
[57] E. Chu, P. Liu, MeanSum: A neural model for unsupervised multi-document abstractive summarization, in: Proceedings of the 36th International Conference on Machine Learning, vol. 97, 2019, pp. 1223–1232.
[58] Y. Suhara, X. Wang, S. Angelidis, W.-C. Tan, OpinionDigest: A simple framework for opinion summarization, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 5789–5798.
[59] V. Rakesh, W. Ding, A. Ahuja, N. Rao, Y. Sun, C.K. Reddy, A sparse topic model for extracting aspect-specific summaries from online reviews, in: Proceedings of the 2018 World Wide Web Conference, 2018, pp. 1573–1582.
[60] X. Wang, Y. Suhara, N. Nuno, Y. Li, J. Li, N. Carmeli, S. Angelidis, E. Kandogann, W.-C. Tan, ExtremeReader: An interactive explorer for customizable and explainable review summarization, in: Companion Proceedings of the Web Conference 2020, 2020, pp. 176–180.
[61] E. Martínez-Cámara, V. Shwartz, I. Gurevych, I. Dagan, Neural disambiguation of causal lexical markers based on context, in: IWCS 2017 — 12th International Conference on Computational Semantics — Short Papers, 2017.
[62] R. Wities, V. Shwartz, G. Stanovsky, M. Adler, O. Shapira, S. Upadhyay, D. Roth, E. Martinez Camara, I. Gurevych, I. Dagan, A consolidated open knowledge representation for multiple texts, in: Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-Level Semantics, 2017, pp. 12–24.
[63] R. Agrawal, T. Imieliński, A. Swami, Mining association rules between sets of items in large databases, SIGMOD Rec. 22 (2) (1993) 207–216.
[64] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1) (1977).
[65] G. Ganu, N. Elhadad, A. Marian, Beyond the stars: Improving rating predictions using review text content, in: WebDB, 2009.
[66] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 1967, pp. 281–297.
[67] J. Jagarlamudi, H. Daumé III, R. Udupa, Incorporating lexical priors into topic models, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012, pp. 204–213.
[68] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13, Curran Associates Inc., 2013, pp. 3111–3119.
[69] C.J. Carmona, M.J. del Jesus, F. Herrera, A unifying analysis for the supervised descriptive rule discovery via the weighted relative accuracy, Knowl.-Based Syst. 139 (2018) 89–100.
[70] G. Dong, J. Bailey, Contrast Data Mining: Concepts, Algorithms, and Applications, first ed., Chapman & Hall/CRC, 2012.