Expert Systems With Applications 209 (2022) 118294

Contents lists available at ScienceDirect

Expert Systems With Applications

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

An end-to-end ranking system based on customers reviews: Integrating semantic mining and MCDM techniques

Milad Eshkevari [a,1], Mustafa Jahangoshai Rezaee [a,*,2], Morteza Saberi [b,3], Omar K. Hussain [c,4]

[a] Faculty of Industrial Engineering, Urmia University of Technology, Urmia, Iran
[b] School of Information, Systems and Modelling, University of Technology Sydney, NSW, Australia
[c] School of Business, University of New South Wales, Canberra, ACT, Australia

ARTICLE INFO

Keywords:
End-to-end ranking
Aspect-based sentiment analysis
Dawid-skene algorithm
Best-worst method
Customer reviews
Sydney hotels ranking

ABSTRACT

Considering customer reviews is one of the challenges of real-world decision models. These reviews can be on different platforms and include a large amount of information and may also include incomprehensible and unrelated phrases. The main advantage of customer reviews is the realistic view that it provides a realistic view of the product or service. Therefore, converting unstructured and incoherent customer-based reviews into machine learning language and ultimately turning it into a decision model is very important. In this paper, we propose an end-to-end ranking method for integrating mechanisms such as text processing, sentiment analysis and the multi-criteria decision-making technique. The proposed ranking method relies on the integration of three methods, namely, the aspect-based sentiment analysis (ABSA) method, the Dawid-Skene algorithm and the Best Worst Method (BWM). In other words, the proposed work encompasses four major steps: i) crawling customer reviews, ii) preprocessing, iii) aspect term extraction, aspect category detection and polarity detection, and iv) designing a decision-making model. The main contribution of this study is to consider ABSA at three levels simultaneously and integrate ABSA and BWM in designing an end-to-end ranking method for ranking the quality of hotel services, facilities and amenities based on customer reviews. The ability of the proposed end-to-end ranking framework is evaluated using a real data set of user reviews of Sydney hotels.

1. Introduction

Evaluating the quality of service is an important issue for consumers. The hotel industry, which is a hospitality-based service, focuses on improving customer satisfaction by providing outstanding service (Shirouyehzad, Tavakoli, & Badakhshian, 2016). Customer reviews, as a comprehensive reference of customers' voices on the quality of hotel services, contain useful and valuable information. Therefore, customer reviews on the services rendered in a hotel provide useful information on the desirability of staying in that hotel. A hotel's services can be divided into different categories such as food, room, etc. However, due to the high volume of data in the review comments of the hotel users, processing these comments to extract the main information is difficult and tedious. Hence, the development of automated systems to extract the important aspects from user reviews and their polarity has grown.

User-generated content (UGC) is any form of content that has been posted by users on online platforms and includes users' opinions on products or services which reflects their quality ratings. UGC contains review information and is a valuable resource for improving goods and services. However, processing large amounts of unstructured and irregular data is difficult. Many researchers have turned their attention to UGC processing at the document, sentence, and aspect levels. Of these studies, aspect-based sentiment analysis (ABSA) is more complex and requires a great deal of knowledge to extract a user's feeling or opinion about a product or service from the review comments (Tran, Ba, & Huynh, 2019). ABSA is a text analysis technique that categorizes the

* Corresponding author.
E-mail addresses: m.eshkevari@ine.uut.ac.ir (M. Eshkevari), m.jahangoshai@uut.ac.ir (M. Jahangoshai Rezaee), Morteza.Saberi@uts.edu.au (M. Saberi), O.hussain@adfa.edu.au (O.K. Hussain).
[1] Orcid: 0000-0002-8622-6734.
[2] Orcid: 0000-0002-3340-666X.
[3] Orcid: 0000-0002-5168-2078.
[4] Orcid: 0000-0002-5738-6560.

https://doi.org/10.1016/j.eswa.2022.118294
Received 28 October 2021; Received in revised form 27 May 2022; Accepted 25 July 2022
Available online 28 July 2022

sentiment being expressed in data such as customer feedback into domain-specific aspect categories and sentiment polarity (García-Pablos, Cuadros, & Rigau, 2018).

In recent years, ABSA has received special attention at an important event on natural language processing on semantic evaluation tasks, namely SemEval (Pontiki et al., 2016). This event provided associated data from customer reviews in several languages in the areas of restaurants and laptops. In general, ABSA can be intended in three levels (Al-Smadi, Qawasmeh, Al-Ayyoub, Jararweh, & Gupta, 2018): i) opinion target expression, ii) aspect sentiment polarity and iii) aspect category detection. In recent years, ABSA has attracted the attention of many researchers who have studied the issue of ABSA at one or more levels using standard data provided or crawling user reviews from sites such as tripadvisor.com or booking.com (Pontiki et al., 2016). Based on our best knowledge, ABSA hotel reviews have focused on these three levels. In other words, each researcher has only tried to use sentiment analysis to identify users' opinions about aspects, aspect categories or the polarity of their opinion separately, but they have not used this extracted information to build a useful decision-making tool based on real customer reviews. Where the ultimate goal of information processing and analysis methods is to turn them into a comprehensive decision or opinion. Therefore, there is a need for a framework to use ABSA in making practical decisions and opinions.

To deal with this drawback, this study contributes to the existing literature by integrating three levels of ABSA and using them for constructing an end-to-end decision-making model. In this paper, an end-to-end hotel ranking approach is proposed integrating ABSA and the multi-criteria decision-making method. Due to the conflicting opinions of reviewers, these systems have a complex structure, including several systems and related subsystems. As a result, ranking units in these systems is very difficult and requires the use and integration of different methods and algorithms. To remove these difficulties and improve the ranking performance, end-to-end analysis frameworks have been developed. In the proposed method, ABSA is used to identify the polarity of the users' opinions on each of the aspects relating to hotel quality. First, using the rule-based method, the aspect categories in each of the user reviews are extracted. Second, using the machine learning model, the polarity of the users' opinions on each of the identified aspect categories is determined. Third, using the Dawid-Skene algorithm, the users' opinions on each aspect are aggregated and considered as input to the Best-Worst Method (BWM) as a well-known and powerful decision-making method. Finally, BWM determines the ranking of each hotel using the values obtained from the aggregation of user opinions on each of the aspects (as the decision criteria). The proposed approach combines ABSA's ability to extract useful information from raw user reviews and the BWM method and provides a powerful and fact-based platform for building an appropriate hotel decision-making system that can be a good and appropriate solution for hotel service introduction or promotion sites. While the networked environment empowers tourists by enabling them to access the vast number of accommodation providers, these tourists may have a difficult time making a decision when faced with many alternatives. This issue is termed "choice overload" or "customer fatigue". This fatigue may become more serious if tourists become obsessed with reading other tourists' comments about particular accommodation. Moreover, most services that manage the huge amount of unstructured information on the web are designed to support service providers rather than tourists. In this paper, we propose an end-to-end ranking method for shortlisting the most suitable accommodation for a given tourist. The main contributions of this paper are as follows:

1) Considering ABSA at three levels simultaneously and using it to rank hotels based on customer reviews.
2) Integrating ABSA and BWM in designing an end-to-end ranking method for ranking the quality of hotel services, facilities and amenities based on customer reviews.
3) Supporting tourists to find the most suitable accommodation by addressing the issue of "choice overload".
4) Designing an end-to-end ranking method as a service for web users.

The rest of this paper is organized as follows: Section 2 discusses the relevant literature on review-based sentiment analysis. In Section 3, a brief description of the ABSA algorithm and BWM method is given. The problem statement and the proposed framework for designing an end-to-end customer review-based ranking are given in Section 4. In Section 5, the proposed methodology is implemented through a case study of TripAdvisor's online reviews of Sydney's hotels. Also, the results and related analyses are presented in this section. Theoretical and practical implications and user-centered based evaluation are provided in Sections 6 and 7 respectively. Finally, the summary and conclusion of the paper are given in Section 8.

2. Literature review

In recent years, a lot of research has been conducted on the sentiment analysis of hotel reviews. As previously mentioned, of the various sentiment analysis approaches, ABSA is more complex due to the extraction of segregated information that conforms to different aspects of customer reviews. The complexity and difficulty of this approach, along with its application and attractiveness, have attracted the attention of many researchers (Schouten & Frasincar, 2015). ABSA for hotel reviews, as one of the main SemEval-2016 tasks, has been a natural language processing challenge in recent years and has been pursued by many researchers. Mostly, each of the studies focused on only one of the subtasks including Aspect Term Extraction, Aspect Category Detection, and Polarity Detection.

One of the first works in this area was by Chiu et al. (2005). They developed an e-Negotiation support system with a *meta*-modeling approach in the Web services area. One of the weaknesses of their proposed approach is that it mainly focuses on modelling aspects. Yu et al. (2007) presented algorithms based on end-to-end QoS constraints for Web service selection. Xu et al. (2011) proposed a model to visualize comparative relations between products from customer reviews on Amazon. According to the existence of interdependencies among relations to help companies for determining potential risks. They suggested the outcome of risks identified for designing new products and marketing strategies. Also, Huang et al. (2018) investigated the role of review font on the effect of consumers' perceived reviewer credibility. By combining two methods, conditional random field (CRF) and Latent Dirichlet Allocation (LDA), Laddha & Mukherjee (2018) proposed a method to respond to all three subtasks of Aspect Identification, Aspect Polarity Identification, and Aspect Polarity Identification. In this hybrid method, terms are first extracted from sentences using CRF. Then, using LDA, each of these terms is assigned to one of the predefined categories. Then, in the form of a regression approach, the degree of polarity of users' opinions is predicted.

Using a combination of the lexicalized domain ontology and the neural network rotatory attention mechanism (LCR-Rot), Wallaart & Frasincar (2019) developed a two-step method for detecting the polarity of user (customer) reviews on hotels. In this paper, using the lexicalized domain ontology, the aspects are extracted and then using a rotatory attention mechanism in the neural network, the polarity of each is determined. Zhang, Li & Song (2019) developed the Aspect-specific Graph Convolutional Network (GCN) method for considering syntactical constraints and long-range word dependencies in the ABSA of hotel reviews. Also, Li et al. (2019) (recognized the polarization of opinions in the ABSA of hotel reviews by restructuring the Attention-based bi-directional Long Short term Memory (BiLSTM- Attention) model and sharing in-layer information. Wang et al. (2019) proposed a supply chain involving a manufacturer and a retailer. This supply chain used customer reviews. In this structure, the customers purchase the product and provide the review about their own decision, the manufacturer

Table 1
A summary of the reviewed related state-of-the-art works.

Reference	Sub-task1 (S1)	Sub-task2 (S2)	Sub-task3 (S3)	Integrating Sub-tasks in decision making	Model	Accuracy
(Laddha & Mukherjee, 2018)	✓	✓	✓	✗	(S1, S2, S3):LDA + CRF	0.65
(Wallaart & Frasincar, 2019)	✗	✗	✓	✗	S3:LCR-Rot	0.88
(Movahedi, Ghadery, Faili, & Shakery, 2019)	✗	✓	✗	✗	S2:TAN	0.82
(Zhang, Li, & Song, 2019)	✗	✗	✓	✗	S3:GCN	0.79
(Li et al., 2019)	✗	✗	✓	✗	S3:BiLSTM-Attention	0.83
(Li, Bing, Zhang, & Lam, 2019)	✓	✓	✓	✗	(S1, S2, S3):BERT + SAN	0.75
(Chen et al., 2020)	✗	✗	✓	✗	S3:CoGAN	0.87
(Truşcă, Wassenberg, Frasincar, & Dekker, 2020)	✗	✗	✓	✗	S3:Deep Contextual Word Embeddings + Hierarchical Attention	0.81
(Zhang, Xu, & Zhao, 2020)	✗	✗	✓	✗	S3:CMA-MemNet	0.81
(Liu & Shen, 2020)	✗	✗	✓	✗	S3:GANN	0.80
(Karimi, Rossi, & Prati, 2020)	✗	✗	✓	✗	S3:BERT	0.86
(Huang, Meng, Guo, Ji, & Han, 2020)	✓	✗	✓	✗	S1:JASen S3:JASen	0.81 0.84
(Tang, Ji, Li, & Zhou, 2020)	✗	✗	✓	✗	S3:DGEDT-BERT	0.84
Ali et al. (2021)	✗	✓	✗	✗	S2:LDA	–
(Alamoudi and Alghamdi, 2021)	✗	✗	✓	✗	S3:Deep learning	0.83
(Li et al. 2021)	✓	✗	✓	✗	(S1,S3): JTSG	0.72
(Kumar et al. 2021)	✓	✗	✓	✗	(S1,S3):Bi-LSTM-CRF	0.78
(Wang et al. 2021)	✗	✗	✓	✗	S3:Deep learning	0.86
Our framework	✓	✓	✓	✓	S1:Rule-based S2:Rule-based S3:Pre-processing + LR + different features	100 100 0.88

decides about quality and price, and the retailer determines the own price. For this purpose, they presented the Stackelberg game and found equilibrium for the supply chain.

Movahedi, Ghadery, Faili, & Shakery (2019) used a bi-directional Gated recurrent unit (GRU) and introduced the Topic-Attention Network (TAN) model to deal with aspect category detection. This paper introduced the Joint Aspect-Sentiment Topic Embedding model which performed two subtask aspect term extractions and polarity detections simultaneously. Combining two models, Bidirectional Encoder Representations from Transformers (BERT) (Devlin, Chang, Lee, & Toutanova, 2018) and Self-Attention Networks (SAN), Li, Bing, Zhang, & Lam (2019) examined three ABSA subtasks on hotel reviews simultaneously in the form of a sequence labelling problem.

Using the Cooperative Graph Attention Networks (CoGAN) model, Chen et al. (2020) used the document-level sentiment preference modeling approach to determine the polarity of opinions in the ABSA of hotel reviews. Recently, Truşcă, Wassenberg, Frasincar, & Dekker (2020) used a combination of Deep Contextual Word Embeddings and Hierarchical Attention to determine the polarity of opinions in ABSA on hotel reviews. Liu & Shen (2020) introduced the Gated Alternate Neural Network (GANN) to determine the polarity of opinions in the ABSA of hotel reviews. Also, Zhang, Xu, & Zhao (2020) introduced a new method called the convolutional multi-head self-attention memory network (CMA-MemNet) to detect the polarity of opinions in the ABSA of hotel reviews. In this regard, by changing the structure of the BERT model and fine-tuning it, Karimi, Rossi, & Prati (2020) developed a new model for detecting the polarity of opinions in the ABSA of hotel reviews. Finally, Tang, Ji, Li, & Zhou (2020) introduced a dependency graph enhanced dual-transformer network (DGEDT) and combined it with the BERT method to present another approach to detect the polarity of the opinions in the ABSA of hotel reviews.

Arenas-Márquez et al. (2021) used an encoding algorithm based on a convolutional neural network (CNN) for online reviews of travel group-type topics extracted from TripAdvisor. They showed that CNNs encoding can be used for identifying the unique topics associated with the data structure determined by the four trip types. Zhang et al. (2021) proposed online customer reviews for mining product innovation ideas. For this purpose, they used an ensemble method for multiple word embeddings and applied a novel deep learning approach with the focal loss function to handle the imbalanced classification problem. They used the proposed approach on a dataset of 10,000 customer reviews from Amazon and achieved the Area Under Curve (AUC) score of 0.91 and an F1-score of 0.89.

Xu (2021) examined the effect of positive and negative closed-form evaluations on customer online review behavior. They found the open-ended comments better reflect overall customer satisfaction than the closed-form evaluations. In contrast, the closed-form evaluations have more significant reflect customers' low overall satisfaction. Finally, Kim and Lim (2021) proposed a method based on integrating sentiment and statistical process control (SPC) analyses. In the method, the customer satisfaction score is determined by the sentiment analysis from customer review data. Also, the SPC chart analysis helps early detection of customer complaints and prevents possible failures.

Ali et al. (2021) provided a new technique by combining Latent Dirichlet Allocation (LDA) and lexicon-based algorithms to gain insights into Marrakech's e-reputation, enhancing the tourism experience in this city. Alamoudi and Alghamdi (2021) presented aspect-based sentiment analysis on yelp reviews using deep learning and word embedding. This study was conducted with aspect polarity detection and compared results with machine learning models. Li et al. (2021) presented a novel end-to-end generative model based on encoder-decoder, namely Joint Term-Sentiment Generator (JTSG), which is presented to conduct aspect term extraction and aspect polarity detection simultaneously. Kumar et al. (2021) presented a framework for detecting aspects and their polarities simultaneously using Bi-LSTM hybridized with CRF. Wang et al. (2021) proposed a novel network with multiple attention mechanisms for aspect polarity detection using to combine BERT model and multiple attention mechanisms, including intra- and inter-level attention mech.

The main reviewed state-of-the-art ABSA approaches are summarized in Table 1.

As can be seen, so far, all ABSA studies have been considered at the level of one of the three subtasks, but no attempt has been made to use ABSA to develop a decision-making approach for automated hotel

```xml
<Review rid="1014458">
 - <sentences>
   - <sentence id="1014458:0">
       <text>I have eaten at Saul, many times, the food is always consistently, outrageously good.</text>
     - <Opinions>
         <Opinion to="42" from="38" polarity="positive" category="FOOD#QUALITY" target="food"/>
       </Opinions>
     </sentence>
   - <sentence id="1014458:1">
       <text>Saul is the best restaurant on Smith Street and in Brooklyn.</text>
     - <Opinions>
         <Opinion to="4" from="0" polarity="positive" category="RESTAURANT#GENERAL" target="Saul"/>
       </Opinions>
     </sentence>
   - <sentence id="1014458:2">
       <text>The duck confit is always amazing and the foie gras terrine with figs was out of this world.</text>
     - <Opinions>
         <Opinion to="69" from="42" polarity="positive" category="FOOD#QUALITY" target="foie gras terrine with figs"/>
         <Opinion to="15" from="4" polarity="positive" category="FOOD#QUALITY" target="duck confit"/>
       </Opinions>
     </sentence>
   - <sentence id="1014458:3">
       <text>The wine list is interesting and has many good values.</text>
     - <Opinions>
         <Opinion to="13" from="4" polarity="positive" category="DRINKS#STYLE_OPTIONS" target="wine list"/>
         <Opinion to="13" from="4" polarity="positive" category="DRINKS#PRICES" target="wine list"/>
       </Opinions>
     </sentence>
   - <sentence id="1014458:4">
       <text>For the price, you cannot eat this well in Manhattan.</text>
     - <Opinions>
         <Opinion to="0" from="0" polarity="positive" category="RESTAURANT#PRICES" target="NULL"/>
         <Opinion to="0" from="0" polarity="positive" category="FOOD#QUALITY" target="NULL"/>
       </Opinions>
     </sentence>
   </sentences>
</Review>
```


Fig. 1. A sample sentence from the SemEval-2016 dataset (Pontiki et al., 2016).

Table 2
SemEval-2016 restaurant reviews dataset description.

Attributes	Information
Number of Reviews	350
Number of Sentences	2000
Number of Aspects	721
Aspect Categories	RESTAURANT#GENERAL, FOOD#QUALITY, RESTAURANT#MISCELLANEOUS, FOOD#PRICES, DRINKS#QUALITY, LOCATION#GENERAL, RESTAURANT#PRICES, AMBIENCE#GENERAL, DRINKS#STYLE_OPTIONS, SERVICE#GENERAL, DRINKS#PRICES, FOOD#STYLE_OPTIONS
Polarities	positive, negative, neutral

ranking. In this study, we introduce, for the first time, the combination of ABSA with multi-criteria decision-making to propose an automated approach based on user reviews and comments for hotel ratings. In this method, ABSA at Level 3 of subtasks and the extraction of aspects in hotel reviews along with the polarity of their opinions are considered as decision criteria for use in the BWM method. In this way, quality ratings based on user reviews can be provided to hotels without interference and with an end-to-end approach. Due to conflicting goals, most machine learning models and responding to users in real-time, ranking systems have a complex structure including several systems and related subsystems. To remove these limitations and improve the performance of the analysis, end-to-end analysis frameworks are developed. It uses different algorithms and methods and every end-to-end ranking system. A review of these frameworks for production quality ranking can be found in Iqbal et al. (2019). Singh and Lee (2016) proposed an end-to-end framework based on convolutional neural networks (CNNs) to simultaneously localize and rank relative visual attributes. They examined the proposed framework on various datasets and showed it is much faster than previous methods. Wu et al. (2018) proposed an end-to-end neural matching framework for an E-commerce sponsored search and optimized the pointwise cross-entropy loss. To validate the performance of the proposed framework, they applied it to the real traffic of a large-scale e-commerce sponsored search.

Papadimitriou et al. (2021) analyzed the characteristics of scientific applications using an end-to-end data collection of workflow performance data to better understand their impact on the underlying infrastructure. To validate the results, they considered two classes of real-world workflows including the I/O-intensive genome analysis and the CPU and memory-intensive material science workflows. Yu et al. (2021) used end-to-end text detection in videos with online tracking. For this purpose, they considered an explainable descriptor, combining appearance, geometry and PHOC features, to build an end-to-end video text detection with online tracking. The results show an improvement in the F-score of more than 2 % and 81.52 % on the Minetto dataset. Other successful end-to-end approaches can be found in Feichtenhofer et al. (2017) for object tracking, Liu et al. (2018) for text spotting and Xu et al. (2019) for zero-shot cross-modal retrieval.

3. Methodologies

In this section, the methodologies used in this study are briefly introduced. First, the notations used in this study are listed below.

- C : Set of evaluation criteria
- A_B: Priority of important criterion over the other criteria
- A_W: Priority of the other criteria based on the worst criterion
- N: Number of customer reviewers
- M: Number of hotels
- M_p: Number of hotels in the star category
- N_{Mp}: Number of reviews per hotel in the star category
- K: Number of aspect terms
- P: Number of aspect categories
- ε : Accuracy of the learning algorithm in the Dawid-Skene algorithm
- γ : Clustering parameter in the Dawid-Skene algorithm

3.1. Aspect-Based sentiment analysis

ABSA analyzes the sentiments in the content according to specific and important aspects. In other words, ABSA processes sentiments in the content by its aspects. ABSA is generally divided into the following four tasks (Al-Smadi, Talafha, Al-Ayyoub, & Jararweh, 2019): 1) aspect-term extraction, 2) aspect-term polarity identification, 3) aspect-category identification, and 4) aspect-category polarity.

On the other hand, in a general view, ABSA can be performed in three subtasks)Al-Smadi, Qawasmeh, Al-Ayyoub, Jararweh, & Gupta, 2018): 1) opinion target expression, 2) aspect sentiment polarity and 3) aspect category detection. In Subtask 1, the aspect terms in the content are

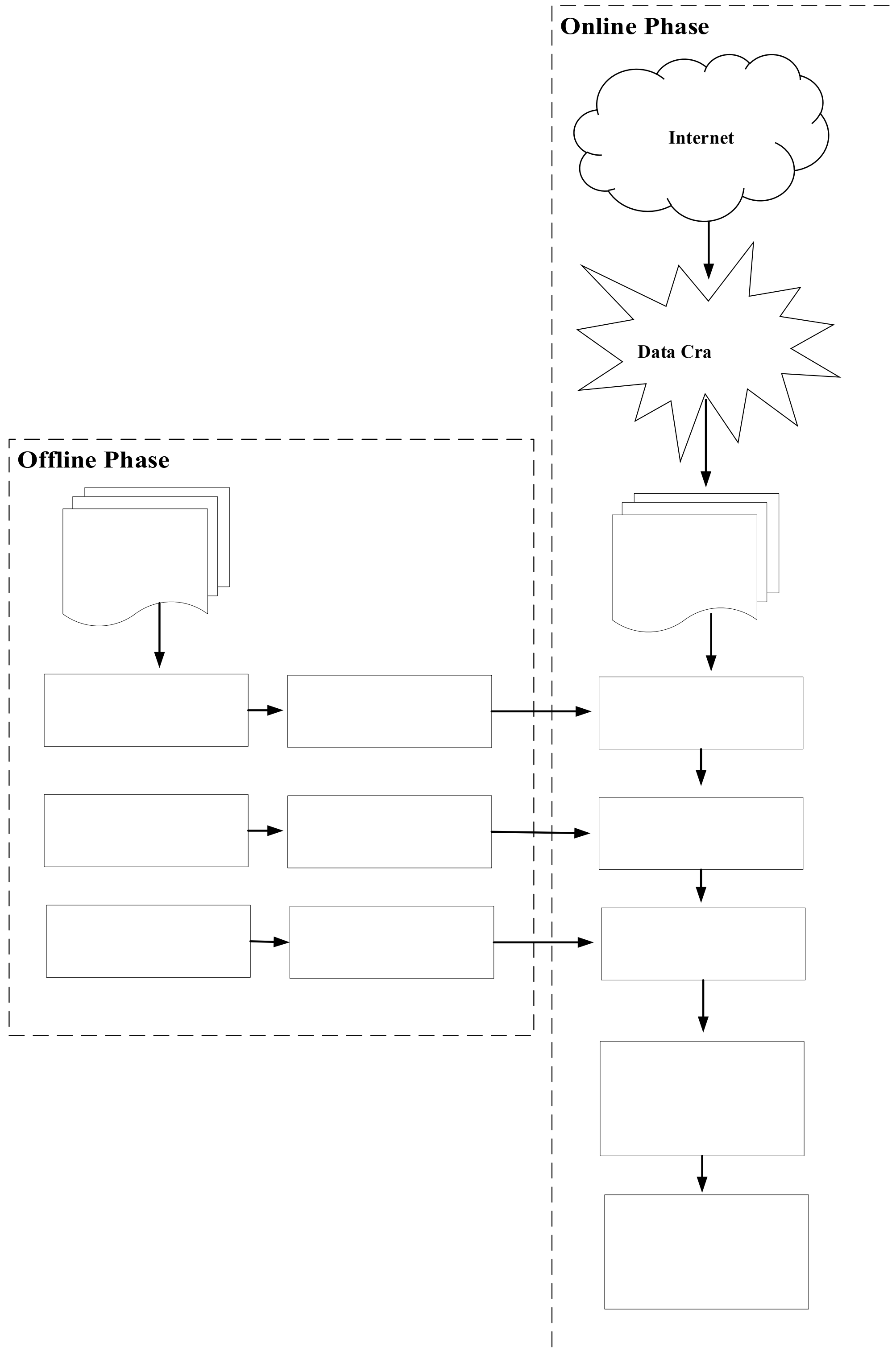

Fig. 2. An overview of the proposed framework.

Table 3
First preprocessing in the aspect polarity detection stage.

Before Preprocessing	Judging from previous posts, this used to be a ***good*** place, but not longer
After Preprocessing	Judging from previous posts, this used to be a ***Positive*** place, but not any longer

Table 4
Second preprocessing in the aspect polarity detection stage.

Before Preprocessing	Judging from previous posts this used to be a Positive place, but not any longer
After Preprocessing	Judging from previous posts this used to be a Positive place, but not any***_NOT longer_NOT***

Table 5
Comparison of implemented models in polarity detection.

Model	Accuracy	F_1	Time (second)	Model	Accuracy	F_1	Time (second)
MLP + FF	0.78 ∓ 0.21	0.51 ∓ 0.18	77	LR + FF	0.85 ∓ 0.02	0.66 ∓ 0.01	67
MLP + BF	0.85 ∓ 0.17	0.70 ∓ 0.16	78	LR + BF	0.86 ∓ 0.01	0.69 ∓ 0.01	65
MLP + TF	0.84 ∓ 0.14	0.62 ∓ 0.15	78	LR + TF	0.86 ∓ 0.01	0.66 ∓ 0.01	61
MLP + FF + BF + TF	0.84 ∓ 0.12	0.64 ∓ 0.13	83	**LR + FF + BF + TF**	0.88 ∓ 0.01	0.73 ∓ 0.01	69
KNN + FF	0.64 ∓ 0.09	0.60 ∓ 0.07	97	SVM + FF	0.82 ∓ 0.02	0.48 ∓ 0.02	73
KNN + BF	0.73 ∓ 0.08	0.59 ∓ 0.08	94	SVM + BF	0.81 ∓ 0.02	0.36 ∓ 0.01	72
KNN + TF	0.76 ∓ 0.08	0.62 ∓ 0.07	102	SVM + TF	0.79 ∓ 0.02	0.19 ∓ 0.01	*73*
KNN + FF + BF + TF	0.73 ∓ 0.09	0.59 ∓ 0.07	112	SVM + FF + BF + TF	0.80 ∓ 0.02	0.32 ∓ 0.02	75
LSTM + Monogram Word Embedding	0.84 ∓ 0.15	0.66 ∓ 0.10	237	CNN + Monogram Word Embedding	0.85 ∓ 0.12	0.68 ∓ 0.10	310
LSTM + Bigram Word Embedding	0.85 ∓ 0.13	0.67 ∓ 0.11	245	CNN + Bigram Word Embedding	0.85 ∓ 0.11	0.71 ∓ 0.11	320
LSTM + Trigram Word Embedding	0.85 ∓ 0.12	0.69 ∓ 0.11	256	CNN + Trigram Word Embedding	0.83 ∓ 0.11	0.56 ∓ 0.10	317

Table 6
Wilcoxon signed-rank test.

	FF	BF	TF	FF + BF + TF
Accuracy	0.75	1	0.875	1
F1	1	0.875	0.875	0.875

identified and separated from the other words. In Subtask 2, the identified aspects are grouped into predefined aspect categories. Subtask 3 determines the user's polarity toward the identified aspect(s). The data presented in the SemEval-2016 Task contains restaurant reviews with one or more sentences. Each of these sentences contains users' opinions on one or more aspects. In this dataset, each aspect is specified as a target and represents the word(s) that expresses the user's opinion about a specific aspect category. Also, for each aspect category, a polarity is provided that indicates whether the user's view of that aspect category is positive, negative, or neutral. Fig. 1 shows an example of a SemEval-2016 dataset in XML format (Pontiki et al., 2016).

Table 2 presents the SemEval-2016 dataset information in the restaurant reviews section.

3.2. Best worst method

The best worst method is one of the newest and most effective multi-criteria decision-making methods proposed by Rezaei (2015) and is used to determine weights according to decision criteria. One of the salient features of BWM compared to other multi-attribute decision methods is that it provides stronger comparisons despite the need for less comparative data and more reliable results (Valipour, Yousefi, Rezaee, & Saberi, (2021)). Using only two vectors, instead of a perfect pairwise comparison matrix, BWM requires less comparison than a perfect pairwise comparison matrix and makes the process more consistent between comparisons because comparisons are made with this highly structured

Table 7
Overview of crawled data information.

Hotels	2-Star Hotels	3-Star Hotels	4-Star Hotels	5-Star Hotels
Number of hotels	79	113	96	39
Crawled Reviews	397	596	771	531

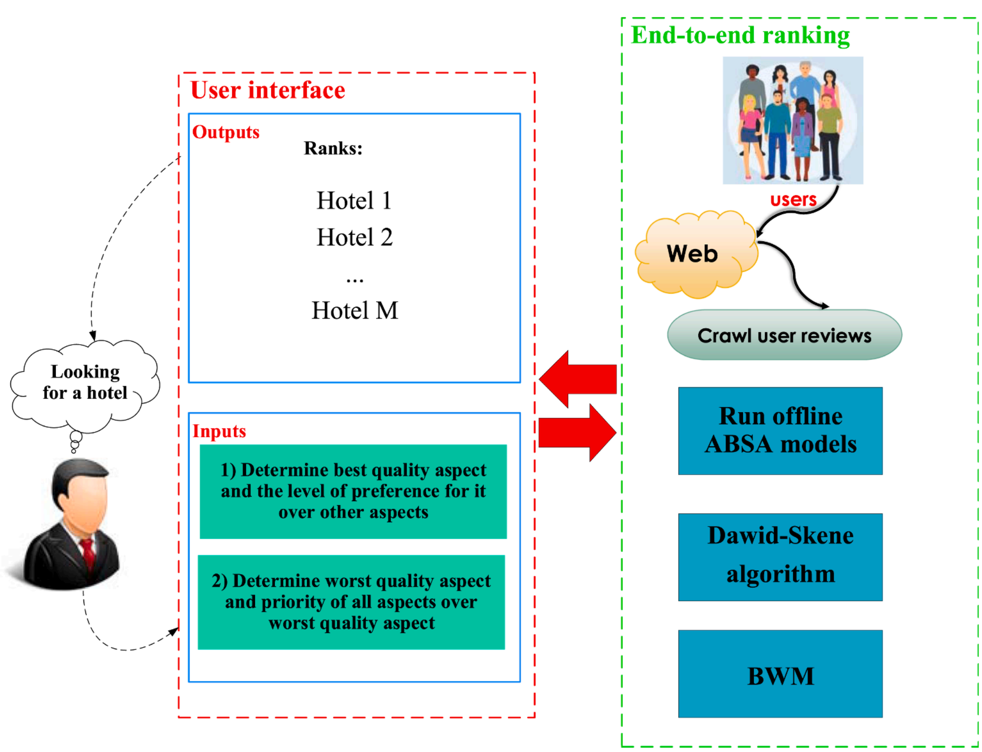

Fig. 3. A schematic view of the proposed framework.

Table 8
A sample of aggregated data for all aspects of different hotels.

Hotels	RESTAURANT#GENERAL	FOOD#QUALITY	...	FOOD#STYLE_OPTIONS
Hotel 1	−1	0.1	...	1
Hotel 2	0.3	0.3	...	0.3
Hotel 3	0.4	−0.2	...	−1
...	...	...	...	...
Hotel N	−0.8	0.4	...	0.7

Table 9
Pairwise comparison vector for the best criterion in 2-star hotels.

Criteria	Best Criteria (RESTAURANT#PRICES)	Criteria	Best Criteria (RESTAURANT#PRICES)
RESTAURANT#GENERAL	2	RESTAURANT#PRICES	1
FOOD#QUALITY	3	AMBIENCE#GENERAL	7
RESTAURANT#MISCELLANEOUS	9	DRINKS#STYLE_OPTIONS	6
FOOD#PRICES	2	SERVICE#GENERAL	5
DRINKS#QUALITY	5	DRINKS#PRICES	3
LOCATION#GENERAL	8	FOOD#STYLE_OPTIONS	4

Table 10
Pairwise comparison vector for the worst criterion in 2-star hotels.

Criteria	**Worst Criteria** (RESTAURANT#MISCELLANEOUS)	Criteria	**Worst Criteria** (RESTAURANT#MISCELLANEOUS)
RESTAURANT#GENERAL	8	RESTAURANT#PRICES	9
FOOD#QUALITY	4	AMBIENCE#GENERAL	3
RESTAURANT#MISCELLANEOUS	1	DRINKS#STYLE_OPTIONS	4
FOOD#PRICES	7	SERVICE#GENERAL	5
DRINKS#QUALITY	5	DRINKS#PRICES	7
LOCATION#GENERAL	2	FOOD#STYLE_OPTIONS	6

Table 11
Pairwise comparison vector for the best criterion in 3-star hotels.

Criteria	Best Criteria (RESTAURANT# GENERAL) (FOOD#QUALITY)	Criteria	**Best Criteria** (RESTAURANT#GENERAL) (FOOD#QUALITY)
RESTAURANT#GENERAL	1	RESTAURANT#PRICES	2
FOOD#QUALITY	1	AMBIENCE#GENERAL	9
RESTAURANT#MISCELLANEOUS	8	DRINKS#STYLE_OPTIONS	6
FOOD#PRICES	3	SERVICE#GENERAL	5
DRINKS#QUALITY	4	DRINKS#PRICES	4
LOCATION#GENERAL	6	FOOD#STYLE_OPTIONS	5

Table 12
Pairwise comparison vector for the worst criterion in 3-star hotels.

Criteria	**Worst Criteria** (AMBIENCE#GENERAL)	Criteria	**Worst Criteria** (AMBIENCE#GENERAL)
RESTAURANT#GENERAL	9	RESTAURANT#PRICES	8
FOOD#QUALITY	9	AMBIENCE#GENERAL	1
RESTAURANT#MISCELLANEOUS	2	DRINKS#STYLE_OPTIONS	4
FOOD#PRICES	7	SERVICE#GENERAL	5
DRINKS#QUALITY	6	DRINKS#PRICES	6
LOCATION#GENERAL	4	FOOD#STYLE_OPTIONS	5

Table 13
Pairwise comparison vector for the best criterion in 4-star hotels.

Criteria	**Best Criteria** (SERVICE#GENERAL)	Criteria	**Best Criteria** (SERVICE#GENERAL)
RESTAURANT#GENERAL	2	RESTAURANT#PRICES	4
FOOD#QUALITY	4	AMBIENCE#GENERAL	5
RESTAURANT#MISCELLANEOUS	6	DRINKS#STYLE_OPTIONS	9
FOOD#PRICES	7	SERVICE#GENERAL	1
DRINKS#QUALITY	7	DRINKS#PRICES	8
LOCATION#GENERAL	3	FOOD#STYLE_OPTIONS	4

Table 14
Pairwise comparison vector for the worst criterion in 4-star hotels.

Criteria	Worst Criteria (DRINKS#STYLE_OPTIONS)	Criteria	Worst Criteria (DRINKS#STYLE_OPTIONS)
RESTAURANT#GENERAL	8	RESTAURANT#PRICES	6
FOOD#QUALITY	6	AMBIENCE#GENERAL	5
RESTAURANT#MISCELLANEOUS	4	DRINKS#STYLE_OPTIONS	1
FOOD#PRICES	3	SERVICE#GENERAL	9
DRINKS#QUALITY	3	DRINKS#PRICES	2
LOCATION#GENERAL	7	FOOD#STYLE_OPTIONS	6

Table 15
Pairwise comparison vector for the best criterion in 5-star hotels.

Criteria	Best Criteria (RESTAURANT#GENERAL)	Criteria	Best Criteria (RESTAURANT#GENERAL)
RESTAURANT#GENERAL	1	RESTAURANT#PRICES	5
FOOD#QUALITY	2	AMBIENCE#GENERAL	3
RESTAURANT#MISCELLANEOUS	3	DRINKS#STYLE_OPTIONS	3
FOOD#PRICES	5	SERVICE#GENERAL	2
DRINKS#QUALITY	2	DRINKS#PRICES	6
LOCATION#GENERAL	3	FOOD#STYLE_OPTIONS	3

Table 16
Pairwise comparison vector for the worst criterion in 5-star hotels.

Criteria	Worst Criteria (DRINKS#PRICES)	Criteria	Worst Criteria (DRINKS#PRICES)
RESTAURANT#GENERAL	6	RESTAURANT#PRICES	2
FOOD#QUALITY	5	AMBIENCE#GENERAL	3
RESTAURANT#MISCELLANEOUS	3	DRINKS#STYLE_OPTIONS	3
FOOD#PRICES	2	SERVICE#GENERAL	5
DRINKS#QUALITY	5	DRINKS#PRICES	1
LOCATION#GENERAL	3	FOOD#STYLE_OPTIONS	3

method. Therefore, pairwise comparison problems are reduced using BWM. Another advantage of this method is that BWM uses a highly structured and understandable method to collect the data needed for pairwise comparisons, which leads to very reliable results that can be explained by the decision-maker. It can easily be revised to increase consistency (Aboutorab, Saberi, Asadabadi, Hussain, & Chang, 2018). The steps of this method are as follows (Rezaei, 2015):

Step 1- Determining a set of evaluation criteria: In the first step, a set of evaluation criteria $\{C_1, C_2, \cdots, C_3\}$ is considered that should be used to evaluate alternative options.

Step 2- Determining the best (most important) and worst (least important) criteria: In this step, the decision-maker should generally consider the best criterion and the worst criterion.

Step 3- Determining the preference of the best criterion over the other criteria: In this step, the decision-maker is required to prioritize the most important criterion over the other criteria and assign a value between 1 and 9, where 1 indicates equal importance and 9 means that the best criterion is much more important than the desired criterion. As a result, the best vector for the others (A_B) is determined using Equation (4).

$$A_B = (a_{B1}, a_{B2}, \ldots, a_{Bn}) \tag{4}$$

Step 4- Prioritizing the criteria over the worst criteria: The decision-maker should use a value between 1 and 9 to indicate the priority of the other criteria based on the selected worst criterion, where 1 indicates equal importance and 9 means that the worst criterion is much more important than the desired criterion. The least important criterion, which leads to a vector worse than the others, is denoted by A_W and expressed by Equation (5).

$$A_W = (a_{W1}, a_{W2}, \ldots, a_{Wn}) \tag{5}$$

Step 5- Finding the optimal weights: Finally, the optimal weights are determined. Two different techniques have been proposed for BWM. The first one, proposed by Rezaei in 2015, can lead to several optimal solutions, while the second technique, also presented by Rezaei in 2016, was designed to find unique weights.

4. Problem statement and proposed framework

The main advantage of the end-to-end ranking of hotels over experts or intermediary companies is based on the realistic views of tourists or customers. In the proposed end-to-end ranking approach, by combining mechanisms such as text processing, sentiment analysis, and the multi-criteria decision making (MCDM) technique, a comprehensive approach is proposed to convert unstructured and incoherent user-based reviews into a user-friendly ranking.

The main contributions of this study are summarized as follows. In contrast to previous studies, ABSA at three levels is simultaneously considered and implemented to rank hotels based on customer reviews. Also, ABSA and BWM are integrated (without third-party intervention) in designing an end-to-end ranking method for ranking the quality of hotel services, facilities and amenities based on customer reviews. It can provide an end-to-end ranking method as a service for web users. Moreover, this system may support tourists to find the most suitable accommodation by addressing the issue of "choice overload".

The proposed framework relies on two integrated modules, namely, offline and online modules. In the offline module, using the data from the SemEval-2016 Task (Pontiki et al., 2016), the models identify aspects in the customer reviews (Model 1), determine the category of each aspect (Model 2) and finally identify the polarity of each user review designed for each category (Model 3). In the online module, by crawling the hotel reviews available on platforms such as trip advisor, the content of raw user comments about each hotel is extracted. After implementing their trained models in the offline module on the crawled reviews, the polarities of the users' opinions on the quality of the 12 aspect categories are determined. To feed this into the MCDM-based ranking model, this information should be fused. We employ the Dawid-Skene algorithm (Dawid and Skene, 1979) and aggregate different users' opinions on each aspect category for each hotel. By retrieving, categorizing and aggregating the associated information on the web, the ranking model can be applied. In this work, by using the BWM, a ranking is presented for the different hotels. Fig. 2 shows the structure of the proposed framework.

As shown in Fig. 2, the ABSA can be divided into three steps: Aspect term extraction, Aspect category detection and Polarity detection. Most previous studies have focused only on polarity detection. In contrast to other studies, this research first considers three different ABSA steps simultaneously and then uses the ABSA output to construct end-to-end decision-making and ranking model. Due to the nature of the first two steps including aspect term extraction and aspect category detection and because of the low accuracy of the previous methods (maximum between 0.5 and 0.6), we use the rule-based approach to achieve the highest accuracy in these two steps. By increasing the accuracy of the first two steps, the result of the third step will improve. The result of this study approach in the Polarity Detection step is also compared with state-of-the-art methods. Finally, a decision-making model for end-to-end hotel ranking is presented that has not been addressed in previous studies.

In the following, the details of the two modules, online and offline, are described. The pseudo-code of the proposed end-to-end ranking

Table 17
Prioritizing the 2-star hotels using BWM concerning ABSA of crawled reviews.

Hotel Name	Score	Rank	Hotel Name	Score	Rank
Wake Up! Sydney	0.72	36	Westend Backpackers	0.63	68
Bondi Backpackers	0.73	35	Criterion Hotel Sydney	0.71	38
Blue Parrot Backpackers	0.76	21	Mad Monkey Backpackers Kings Cross	0.71	39
The Pod Sydney	0.76	22	Elephant Backpacker	0.77	20
Cambridge Lodge Guest House	0.71	40	The Village Bondi Beach - Hostel	0.74	34
The Jensen Potts Point	0.67	56	Holiday Lodge Hotel	0.67	57
Sydney Central Inn	0.67	55	Central Private Hotel	0.85	7
790 on George	0.80	13	Sydney Space	0.65	67
Noah's Bondi	0.67	59	Manly Beach House	0.84	9
Cityview Studio Accommodation	0.80	12	Newtown Space	0.67	58
Billabong Gardens	0.67	54	World Square Hostel	0.74	33
Nates Place Backpackers	0.77	16	The Capsule Hotel	0.67	60
O'Malley's Hotel	0.67	53	Bayswater Boutique Lodge	0.67	52
Backpackers HQ	0.60	74	Sandy Bottoms Guesthouse	0.77	19
Kingsview	0.50	78	Space Q Capsule Hotel	0.67	51
Sydney Star Backpackers	0.85	6	Manly Beachside Apartments	0.93	2
Sydney Central Backpackers	0.67	50	Wynyard Hotel	0.80	14
Funk House Backpackers Hostel Sydney	0.71	41	Coogee Beach House	0.74	32
Glebe Village Backpackers	0.69	47	The Downing Hostel	0.74	31
Home Backpackers	0.76	23	Kings Cross Backpackers	0.67	49
Asylum Sydney Backpackers Hostel	0.94	1	Hump Backpackers	0.56	76
The King's Hotel Sydney	0.77	17	East Sydney Hotel	0.84	8

(continued on next page)

Table 17 (*continued*)

Coogee Beachside Budget Accommodation	0.71	42	Woolbrokers Hotel	0.83	10
Ady's Place Backpackers	0.63	69	Darlo Bar Darlinghurst	0.67	48
Sydney Backpackers	0.63	70	Dunkirk Hotel	0.74	27
Secret Garden Backpackers, Sydney	0.79	15	Sydney Park Hotel	0.87	3
Rooftop Travellers Lodge	0.67	63	Zing Backpackers	0.63	71
Beach Road Hotel	0.48	79	The Original Backpackers Hostel	0.71	44
Sydney Darling Harbour Hotel	0.71	43	Jolly Swagman Backpackers	0.57	75
The Village Surry Hills	0.77	18	Summer House Backpackers	0.67	61
Tiger Hostel Sydney	0.67	62	Ryals Hotel - Broadway	0.74	25
Sinclair's of Surry Hills Budget Accommodation	0.60	73	Casa Central Accommodation	0.76	24
Sydney City Hostel	0.67	64	Mad Monkey Coogee Beach	0.74	26
The Village Broadway	0.61	72	Mad Monkey Backpackers on Broadway	0.85	5
Boardrider Backpacker and Budget Motel	0.74	28	The Village Kings Cross	0.74	30
Glebe Space	0.71	45	The Mercantile Hotel	0.67	65
Sydney University Village	0.87	4	Manly Bunkhouse	0.83	11
Australian Sunrise Lodge	0.53	77	The Australian Heritage Hotel	0.67	66
Manly Beach Hotel Est. 1964	0.74	29	Big Hostel	0.69	46
Balmain Backpackers	0.71	37			

framework is explained as follows.

Algorithm 1 End-to-end hotel quality ranking framework

Offline phase

Inputs:

A set of units' records as ordered pairs (units, raw texts), (units, aspect terms) and (aspect terms, aspect categories) using SemEval-2016 collection

Procedure:

Step1- aspect term extraction by rule-based method:

For i, j **in** (*M* (units), *N* (raw texts)):

For k **in** *K* (aspect terms):

If *k*th aspect in (units(i), raw texts(j))

(units(i), aspect terms(k)) ← 1

End for

End for

Step 2- aspect category detection by rule-based method:

For i **in** *K* (aspect terms):

For j **in** *P* (aspect categories):

If aspect terms (i) in aspect categories (j)

(aspect terms(i), aspect categories (j)) ← 1

End for

End for

Step 3- determining polarity by machine learning model:

For i,j **in** (*M*(units), *N*(processed sentences)):

Extract the sense of each sentence

if sense is positive:

(units (i), sentences (j)) ← 1

else if sense is negative:

(units (i), sentences (j)) ← -1

else (units (i), sentences (j)) ← 0

End for

Outputs:

Database including (units, aspect terms) (Model 1), (aspect terms, aspect categories) (Model 2) and (units, sentences(polarity)) (Model 3)

Online phase:

Inputs:

Outputs of the offline stage.

(*continued on next column*)

(*continued*)

Procedure:

Step 1- crawl hotel reviews from online frameworks

Step 2- categories crawled hotel reviews with regard to hotel stars

Step 3- extract aspect terms, determine aspect categories and polarities by tuned in offline phase

for each star category (1,…, *P*)

for hotels in the star category (1,…, M_p)

for reviews per each hotel in the star category (1,…, N_{Mp})

Extract aspect terms with model 1, determine the aspect categories with model 2 and, determine polarities with regard to each aspect category with model 3.

End for

End for

End for

Step 4- Dawid-Skene algorithm:

Aggregating users' opinions:

each aspect category per each hotel in each star category

Repeat Steps 1 to 4 of the online phase per desired periods

Outputs:

Aggregated polarities of all users' opinions for quality aspects of different hotels

MCDM phase:

Inputs:

Outputs of online stage and user preferences for aspect comparison

Procedure:

Step 1- determining the best and worst aspects by the user (who searches for a hotel)

Step 2- determining the preference of the best aspect over other aspects by user

Step 3- determining the preference of the worst aspect over other aspects by user

Step 4- finding the optimal weights

Outputs:

Ranking all hotels in each star category by BWM

Suppose *N* is the number of customer reviewers and *M* is the number of hotels that be registered on tripadvisor.com. Also, denote *K* and *P* as the number of aspect terms and aspect categories respectively. In the offline phase, the complexity of the algorithm in three steps is$O(MNK)$, $O(PK)$ and $O(MN)$ respectively. Given that$M$ and $N >>> K$ and P, the overall complexity of the algorithm in the first phase is equal to$O(MN)$. Also, in the online phase, assume that in the worst case, all customer reviews for all hotels include whole aspects. Therefore, the complexity of the algorithm in crawled data steps is$O(MNK)$. Also, according to Agarwal, Mandal, Parkes, & Shah (2020), the complexity of the Dawid-Skene algorithm for the clustering phase and learning phase considering γ and ε as the clustering parameter and accuracy of the learning algorithm, are $O(\frac{MN^2}{\gamma^2})$ and $O(\frac{MN^7}{\varepsilon^2})$ respectively. As can be seen, the complexity of the proposed algorithm is dependent on the Dawid-Skene algorithm.

4.1. Offline module

In the offline module, suitable models are designed and identified to perform three different ABSA subtasks. These subtasks are introduced briefly as follows:

Task 1: Design a rule-based method for detecting aspect terms from raw texts using the aspects in the SemEval-2016 restaurant ABSA dataset.

Task 2: Design a rule-based method for assigning each aspect term to the corresponding aspect category concerning the SemEval-2016 restaurant ABSA dataset.

Task 3: Use machine learning methods to determine polarities.

ABSA can be executed by the three approaches (Wallaart & Frasincar, 2019), namely, knowledge-based, machine learning-based and

Table 18
Prioritizing the 3-star hotels using BWM concerning ABSA of crawled reviews.

Hotel Name	Score	Rank	Hotel Name	Score	Rank
Hotel Challis	0.67	14	Ballantyne at Mosman	0.78	7
Rendezvous Hotel Sydney The Rocks	0.37	60	Fawlty Towers Hotel	0.59	26
The Russell Boutique Hotel	0.37	61	DD Apartments On Kent Street	0.76	9
Bondi Beach House Accommodation	0.57	32	MyHoYoHo Apartments Haymarket	0.37	79
The Maisonette	0.47	50	Manly Lodge Boutique Hotel	0.37	80
Mrs Banks Hotel	0.77	8	The Grand Hotel	0.55	40
Springfield Lodge	0.37	62	Central Perk Lodge	0.37	81
Sydney Boutique Hotel	0.55	37	Dixon Residences	0.37	82
1831 Boutique Hotel	0.59	22	Marco Polo Motor Inn - Sydney	0.47	53
Travelodge Hotel Sydney Martin Place	0.50	45	Gazebo Hotel	0.31	109
Ibis Sydney King Street Wharf	0.37	63	Oaks Harmony	0.37	83
Hotel Bondi	0.36	102	DeVere Hotel	0.37	84
Hotel Coronation	0.72	11	Ultimate Apartments Bondi Beach	0.37	85
Metro Apartments On Darling Harbour	0.50	48	Liv Apartments Darling Harbour	0.37	86
The Hughenden Boutique Hotel	0.59	23	Frisco Hotel	0.37	87
Hotel Ravesis	0.29	110	Crown Hotel	0.78	5
Morgans Boutique Hotel	0.59	27	The Merchant Hotel	0.55	41
Royal Exhibition Hotel	0.55	38	Coogee Bay Hotel - Pub Style	0.37	88
The Collectionist	0.67	15	The Grand Central Hotel	0.78	6
Posh Hotel	0.57	31	Park Regis City Centre	0.29	111
Glasgow Arms Hotel	0.37	64	Ibis Budget St Peters	0.57	33
Waldorf Sydney Serviced Apartments	0.55	39	Kafnu Alexandria	0.76	10
Darling Street Apartments	0.37	65	Sydney Hotel QVB	0.92	2
Song Hotel Redfern	0.51	44	Alishan International Guest House	0.37	89

(continued on next page)

Table 18 (*continued*)

The Outback Lodge	0.36	103	Waldorf Drummoyne Serviced Apartments	0.39	59
Sydney Lodge Motel	0.69	12	Napoleon on Kent	0.57	34
Muse Private Hotel	1.01	1	The Merton Hotel	0.37	90
Pensione Italia Bed & Breakfast	0.63	20	Glebe Point YHA	0.37	91
Keg & Brew Hotel	0.59	28	Jackaroo Hostel	0.34	105
Manly Beach Holiday & Executive Apartments	0.37	66	The Savoy Double Bay Hotel	0.80	3
Sydney Boutique Lodges	0.79	4	Sydney Hotel CBD	0.37	92
DD Apartments On Sussex Street (Previously Dulcis Domus Apartments On Sussex)	0.41	57	APX Darling Harbour	0.66	18
Excelsior Serviced Apartments	0.37	67	Song Hotel Sydney	0.55	42
Seventeen Elizabeth Bay Road Apartments Sydney	0.37	68	Ibis Sydney World Square	0.37	93
The Grantham	0.37	69	Ibis budget Sydney East	0.37	94
Cooper Lodge Hotel	0.37	70	Nesuto Chippendale Sydney Apartment Hotel	0.59	29
Hotel Harry, an Ascend Hotel Collection Member	0.66	16	Ibis Sydney Darling Harbour	0.33	107
Metro Apartments on King	0.37	71	Manly Paradise Motel & Apartments	0.37	95
MyHoYoHo City Views Apartments	0.43	55	Arts Hotel	0.60	21
Capitol Square Hotel	0.37	72	Railway Square YHA	0.66	17
Annandale Hotel	0.25	113	Sydney Central YHA	0.55	36
George Street Private Hotel	0.34	104	Sydney Harbour Bed and Breakfast	0.68	13
Edgecliff Lodge Motel	0.50	46	Megaboom City Hotel	0.29	112
The Crescent On Bayswater	0.37	73	Altamont Hotel Sydney - by 8Hotels	0.37	96
Harbour Phoenix Serviced Apartments	0.37	74	The Lord Nelson Brewery Hotel	0.37	97
City Crown Motel	0.37	75	Admiral Collingwood Lodge	0.37	98
Darlinghurst Apartments	0.37	76	Bounce Sydney	0.50	47
Castle Serviced Apartments	0.37	77	Mariners Court Hotel	0.43	56
Reef Resort Apartments	0.39	58	ValueSuites Green Square	0.55	43
City Lodge	0.48	49	Sydney Harbour YHA	0.37	99

(*continued on next page*)

Table 18 (*continued*)

BreakFree on Clarence	0.47	51	Hotel 59	0.37	100
Southern Cross Hotel	0.37	78	Central Railway Hotel	0.37	101
Hotel Steyne	0.34	106	57 Hotel	0.46	54
Sinclairs Serviced Apartments	0.32	108	Medusa	0.64	19
Coogee Prime Lodge	0.47	52	Nesuto Woolloomooloo Sydney Apartment Hotel	0.59	30
Sydney Wattle Hotel	0.59	24	28 Hotel	0.57	35
Lido Suites	0.59	25			

hybrid methods. According to the state-of-art results presented in Table 1, the recent accuracy reported in Subtasks 1 and 2 are not yet close to the appropriate accuracy values. Therefore, the focus of this research is on leveraging the power of the knowledge-based approach and rule-based method to execute Subtasks 1 and 2, thereby identifying the terms and determining their aspect category. Considering that the learning machine approaches did not work well, we used the rule-based method based on the standard database. The learning machine approach is only used for the third task.

In the rule-based method, the presence and absence of the predefined aspects are checked by creating a comprehensive dictionary consisting of the aspect terms presented in the SemEval-2016 Task dataset. Then, the aspect categories are determined concerning the identified aspect terms. For example:

"I was very disappointed with this restaurant." ⟶ detected aspect term=restaurant and detected aspect category=RESTAURANT#GENERAL

After identifying the aspects and determining their categories, the polarity of the user's opinion on the identified aspect categories is determined. For this purpose, the machine learning approach is used. First, the content in which the identified aspect is located goes through two stages of pre-processing. First, a list of words with positive polarity (such as abound, good and easygoing) and negative (such as abnormal, damage and superficially) is identified and stored in a dictionary. Second, using a rule-based method, if these words are observed, they are replaced by positive or negative equivalents. An example of this process is shown in Table 3.

Then, a list of negative words (such as "no", "not", "none", "nobody", "nothing") is defined in a dictionary and if these words are in the content, as shown in Table 4, the next words in the current sentence of content are changed to the word + "_ NOT" before reaching the "." and other separators.

After pre-processing, the processed sentences (sentences of contentment in which one or more aspects are identified) and their corresponding polarities are given to several different machine learning methods. Then, the model which has the better accuracy is identified and selected as the appropriate model. For this purpose, models such as SVM, MLP, Logistic Regression (LR), CNN and LSTM have been used. Furthermore, various combinations of Frequency Features (FF), Bigram Features (BF) and Trigram Features (Ismail, Saidi, Sayadi, & Benbouzid) have been used to increase the accuracy of the models used.

As above described, machine learning algorithms have been used for determining the polarity of opinions regard to each aspect. For this aim, several machine learning and deep learning models were run on the train section of the SemEval-2016 Task dataset. Then, we evaluate the accuracy of models by implementing that in the test section of the SemEval-2016 Task dataset. In each model, we used different combinations of features as input and repeat train and test for 10 trials for each action. Also, in training each model, different structures and parameters have been used by trial and error and reported the best structure result for each model. Table 5 compares the accuracy of each model in term-termsmeathe n and standard deviation of the metrics in 10 trials (mean + -std). As shown in Table 5, the best accuracy in determining the polarity of sentences was obtained by the logistic regression model with a combination of Frequency Features, Bigram Features and Trigram Features.

For testing the hypothesis of the accuracy of the models towards the change of features, the Wilcoxon signed-rank test has been applied. Obtained results are presented in Table 6 indicating P-values.

Concerning Table 6 and obtained P-values in the 5 % significant level, the null hypothesis (H_0: Differences are not significant) is rejected and all differences are significant. In other words, the accuracy of the models has changed significantly with the change in features.

4.2. Online stage

After configuring the rule-based and machine learning models to identify the aspects and their polarity, in the online phase, these models are used to extract the users' opinions on aspects of each hotel. In this regard, as can be seen in Fig. 3, first, the raw text of the users' reviews on each hotel is crawled from https://www.tripadvisor.com. After performing pre-processing such as correcting spelling mistakes and converting all letters to lowercase letters, the processed sentences are given to the models stored in the online phase, respectively. First, the presence of these aspects in the review is checked by a complete dictionary-based model based on the 721 aspects considered in the SemEval-2016 Task dataset. Then, with the aid of the categories presented in the SemEval-2016 Task dataset, the category corresponding to the identified aspect is determined. After identifying the location of the aspects and their category, the trained machine learning polarity detection model (in the offline phase) is applied to detect the polarity of user opinions toward the aspect categories.

After obtaining the output of the machine learning model and determining the polarity of the user's opinion about each aspect, corresponding to the number of users who have commented on each hotel, the dictionaries of opinions on each aspect category are obtained. Each user may have various opinions on each aspect, these being either positive, negative, neutral or "no comment". Therefore, by aggregating the opinions of different users on each hotel, we have a set of different opinions on each aspect of each hotel. To aggregate the opinions, the Dawid-Skene algorithm is used for each aspect category of each hotel. Then, for each aspect and all hotels, values are obtained that indicate the aggregation of opinions of different users towards that aspect category of the hotel.

In the last step of the online phase, hotels are ranked based on the quality of services using aggregated data and the BWM method. This ranking is based on the real opinions of users (customers) and is done without applying the personal opinions of decision-makers. In the real-world application, in proportion to the passage of time and the addition of new comments by users, at certain intervals of time, the actions performed in the offline phase are performed again and all models and related results are updated. Fig. 3 presents a schematic view of the proposed framework including data crawling, using the offline models, and the ranking using BWM.

5. Results: Real-world application

In this section, the proposed end-to-end ranking framework is evaluated using a real data set. For this purpose, user reviews of Sydney hotels have been crawled from https://www.tripadvisor.com. These reviews are unstructured and incoherent information and must be converted into machine learning language and ultimately turn into a decision model.

The structure of data is such that for each hotel, in addition to information such as stars, the opinions of individual users who have reviewed that hotel are also available. For a better comparison, hotels are categorized by stars (2-star, 3-star, 4-star, and 5-star). Table 7 summarizes the crawled data, including hotel groups, the number of hotels in each group, and the number of crawled comments from each group.

Each user's comments are tokenized as independent sentences and the aspects contained in the sentences are identified by a rule-based approach designed in the offline phase. Then, using the model stored in the offline phase, the polarity of users' reviews on the identified aspects is determined.

Using the first rule-based model, the presence or absence of each aspect term in each review has been monitored. Then, using the second rule-based model, identified aspect terms are mapped to the corresponding aspect category. In the next step, using the polarity detection machine learning model, the positive, negative or neutral opinion of each customer (reviewer) towards the identified aspect category will be calculated. In the presence of an aspect term, the customer's aspect category will be identified with 100 % accuracy, and the customer's polarity with respect to that aspect category will be determined with 88 % accuracy (see Table 5).

Using the Dawid-Skene algorithm, users' opinions on each aspect category are aggregated.

Table 8 shows an example of aggregate opinions for each aspect category of each hotel, according to the Dawid-Skene algorithm in the online phase.

As discussed in the previous section, to perform a BWM ranking, the best and worst criteria (aspect category) must first be determined. Then, the advantages of the best criterion over all other criteria and the advantage of such other criteria over the worst criterion are identified. Determining the best and worst criteria and the relationships between them changes according to the star classification of hotels. For example, RESTAURANT#PRICES and RESTAURANT#MISCELLANEOUS were selected as the best and worst criteria for 2-star hotels. Table 9 shows the preference of the best criterion over the other criteria, which have been determined by experts for 2-star hotels. Also, Table 10 shows the priority of all criteria over the worst criterion for 2-star hotels.

Similarly, the best criteria and the worst criteria and the relative relationships between them for 3-star, 4-star and 5-star hotels are set by experts and presented in Tables 11 to 16.

Logically, the values recorded in Tables 9 to16 are based on a sample preference, and using this approach in the real world, anyone can determine the importance of each aspect by their preference and priority.

Finally, using the data provided, the ranking was undertaken for each group of hotels. Tables 17 to 20 show the ranking results of the 2 to

Table 19
Prioritizing of the 4-star hotels using BWM concerning ABSA of crawled reviews.

Hotel Name	Score	Rank	Hotel Name	Score	Rank
Novotel Sydney on Darling Harbour	1.09	57	Liv Apartments	0.78	80
Little Albion, a Crystalbrook Collection Boutique Hotel	1.96	6	Valentine On George	0.78	81
The Grace Hotel Sydney	1.18	47	Best Western Haven Glebe	0.78	82
Best Western Plus Hotel Stellar	0.78	70	Coogee Bay Hotel - Boutique	1.24	42
Hyde Park Inn	1.33	38	Metro Hotel Marlow Sydney Central	0.89	65
PARKROYAL Darling Harbour Sydney	1.37	36	Sydney City Lodge	1.63	14
West Hotel Sydney, Curio Collection by Hilton	1.62	15	Verona Guest House	1.50	24
Adina Apartment Hotel Coogee Sydney	0.89	64	Wyndham Sydney Suites	1.39	33
Q Station Sydney Harbour National Park Hotel	1.14	51	Annam Apartments	0.78	83
Rydges Sydney Central	0.86	67	Castlereagh Boutique Hotel	2.26	2
Quest Potts Point	0.82	69	Novotel Sydney Manly Pacific	1.50	25
Quest Manly	1.41	29	The Urban Newtown	0.78	84

(continued on next page)

Table 19 (*continued*)

Holiday Inn Potts Point - Sydney	1.70	12	Watsons Bay Boutique Hotel	1.28	40
ADGE Apartment Hotel	0.97	62	Dive Hotel	0.78	85
Regents Court Sydney	0.78	71	Adina Apartment Hotel Sydney Chippendale	0.71	94
Rendezvous Hotel Sydney Central	0.78	72	Mantra on Kent	0.78	86
Crowne Plaza Hotel Coogee Beach - Sydney	1.39	32	The York by Swiss-Belhotel	2.24	3
The Bayswater Sydney	1.14	52	Oaks on Castlereagh	0.83	68
The Sydney Boulevard Hotel	1.50	20	Zara Tower Hotel - Luxury Suites and Apartments	1.50	26
Vibe Hotel Sydney	1.41	31	Bondi 38 Serviced Apartments	2.29	1
APX World Square	0.78	73	Simpsons of Potts Point Hotel	1.16	50
Adina Serviced Apartments Sydney Martin Place	1.01	61	Adina Apartment Hotel Sydney Surry Hills	0.78	87
Holiday Inn Darling Harbour	0.74	92	Bet's B&B Studio	1.02	60
BreakFree on George	1.85	9	Spicers Potts Point	1.24	43
Travelodge Hotel Sydney Wynyard	1.12	56	Holiday Inn Old Sydney	0.96	63
Oaks Goldsbrough Apartments	1.44	28	Veriu Central	0.74	93
Cockatoo Island Heritage Houses	1.50	21	Coogee Sands Hotel & Apartments	0.78	88
Travelodge Hotel Sydney	0.87	66	The Tank Stream Hotel	1.39	34
Oaks Hyde Park Plaza	1.55	18	Macleay Hotel	1.85	10
Manor House Boutique Hotel Sydney	1.50	22	The Ultimo	1.14	53
Quality Apartments Camperdown	0.52	96	Vulcan Hotel Sydney	1.38	35
Shakespeare Hotel	0.78	74	Mantra 2 Bond Street	1.14	54
The Kirketon Hotel	0.78	75	Ovolo 1888 Darling Harbour	1.60	17
AeA The Coogee View	1.16	48	Radisson Hotel and Suites Sydney	2.04	4
'The 150 Apartments' by Apartment Hotel	2.00	5	Adina Apartment Hotel Sydney Darling Harbour	0.78	89
Apartment Hotel East Central	1.16	49	Rydges World Square Sydney Hotel	1.50	27
Atelier Serviced Apartments	0.78	76	Cambridge Hotel Sydney	1.29	39
Rydges Camperdown	1.34	37	Adina Apartment Hotel Sydney Town Hall	1.19	45

(*continued on next page*)

Table 19 (*continued*)

Bridal Falls Cottage	0.78	77	Veriu Camperdown	1.24	44
Botanik Apartment Hotel	0.78	78	Veriu Broadway	1.09	58
Seasons Darling Harbour	1.91	7	The Sebel Manly Beach	0.78	90
Astra Apartments Sydney - Kent Street	1.50	23	Four Points by Sheraton, Sydney Central Park	1.62	16
Balmain Wharf Apartments	1.41	30	Novotel Sydney Central	1.19	46
Atlas Serviced Apartments	1.24	41	Adina Apartment Hotel Sydney Central	1.73	11
Madison Carrington Apartments	0.78	79	Metro Aspire Hotel Sydney	1.14	55
Seasons Harbour Plaza	1.91	8	Novotel Sydney Darling Square	1.03	59
Manly Surfside Apartments	1.68	13	Vibe Hotel Rushcutters Bay Sydney	0.78	91
The Great Southern Hotel Sydney	1.55	19	Mercure Sydney	0.57	95

5-star hotels. To facilitate visual comparison, the ranking is divided into four quarters based on customer reviews, as shown in Tables 17 to 20 using the box chart and displayed in different colors. According to the scores obtained from the BWM method, the hotels are divided into four quarters. The green, yellow, blue, and red colors represent the first to fourth quarters, respectively. For this division, the diagrams presented in Fig. 4 are used, which show how the scores assigned to the hotels of each group are distributed in the form of box diagrams.

2-star hotels: Fig. 4-a shows how the scores applied to the 2-star hotels are distributed and shows the deviation of each quarter from the middle of the scores. In this group, the median scores were equal to 0.71, the minimum value was 0.53, the first quarter was 0.65, the third quarter was 0.78 and the maximum value was 0.87. With these values in mind, the scores assigned to each hotel in this group are broken down into quarters 1 to 4. As shown in Table 17, 13 of 79 2-star hotels are in the first quarter (green color) and have the lowest satisfaction level. 30 hotels are in the second quarter (yellow color) and are in the third satisfaction level. 21 hotels are in the third quarter (blue color) and are in the second satisfaction level, and 15 hotels are in the fourth quarter (red color) and are in the first satisfaction level. The Asylum Sydney Backpackers Hostel and Manly Beachside Apartments are ranked first and second, respectively, and the Kingsview and Beach Road Hotel are ranked 78th and 79th, respectively.

3-star hotels: Fig. 4-b shows the distribution of scores applied to 3-star hotels and the degree of deviation of each quarter from the middle of the scores. For this group, the values of minimum, first quarter, middle, third quarter and maximum are calculated as 0.25, 0.37, 0.41, 0.58 and 0.80, respectively.

Similar to the 2-star hotels, the 3-star hotels are divided into four levels based on scores and comparisons with quarters. As shown in Table 18, 30 of the 113 hotels in the fourth quarter (satisfaction level 1), 26 hotels in the third quarter (satisfaction level 2), 3 hotels in the second quarter (satisfaction level 3) and 54 hotels in the first quarter (satisfaction level 4) were ranked. Of the hotels in this group, the Muse Private Hotel and Kafnu Alexandria are ranked first and second, respectively, and Megaboom City Hotel and Annandale Hotel are ranked 112th and 113th, respectively.

4-star hotels: Fig. 4-c shows the distribution of score values applied to 4-star hotels and the deviation of each quarter from the middle of the scores. For this group, the values of minimum, first quarter, middle, third quarter and maximum are equal to 0.52, 0.78, 1.16, 1.50 and 2.29, respectively.

As shown in Table 19, 19 of the 96 hotels in the fourth quarter (satisfaction level 1), 27 hotels in the third quarter (satisfaction level 2), 28 hotels in the second quarter (satisfaction level 3) and 22 hotels in the first quarter (satisfaction level 4) were ranked. Of the hotels in this group, Bondi 38 Serviced Apartments and Castlereagh Boutique Hotel are ranked first and second, respectively, and Oaks on Castlereagh and Quality Apartments Camperdown received the lowest scores with the rank of 95 and 96, respectively.

5-star hotels: Fig. 4-d shows the distribution of the score values applied to the 5-star hotels. The minimum, first quarter, middle, third and maximum values are 0.47, 0.81, 1.25, 1.48 and 2.48, respectively. As shown in Table 20, 10 of the 39 hotels in the fourth quarter (satisfaction level 1), 9 hotels in the third quarter (satisfaction level 2), 10 hotels in the second quarter (satisfaction level 3) and 10 hotels in the first quarter (satisfaction level 4) were ranked. The Ovolo Woolloomooloo and Meriton Serviced Apartments George Street are ranked first and second, respectively, and the Sofitel Sydney Darling Harbour and Shangri-La Hotel Sydney are ranked 38th and 39th, respectively.

Fig. 4 shows that, of the 4 groups, the scores are scattered from highest to lowest by the 4, 5, 2 and 3-star hotels respectively. Negative skewness is observed for the 3-star hotels, which indicates a higher density of hotels in this group at lower satisfaction levels. On the other hand, for the 2 and 5-star hotels, there is a positive skewness and the density of hotels in these groups is at high levels of satisfaction.

As previously discussed, the median scores obtained by the four groups are 0.71, 0.41, 1.16 and 1.25 respectively, indicating that the order of different levels of the hotel has been successful in obtaining the satisfaction of its customers in the form of 5, 4, 2 and 3. Also, the 5 and 4-star hotels have relatively higher scores than the other two groups. For further comparison, the mean, minimum, and maximum scores assigned to the hotels in each group are shown in Fig. 5.

To test the significance of the differences between the mean, minimum, and maximum scores between the four categories including 2, 3, 4, and 5 stars (Fig. 5), the Wilcoxon signed-rank test has been applied. Obtained results are presented in Table 21 indicating P-values.

Concerning Table 7 and obtained P-values in the 5 % significant level null hypothesis (H_0: Differences are not significant) are rejected and all differences are significant. In other words, the differences in minimum, maximum and average have changed significantly with star levels.

In terms of average scores, the 4 and 5-star hotels achieved higher satisfaction scores than the 2 and 3-star hotels. Also, the 4 and 5-star hotels achieved much better minimum and maximum scores than the 2 and 3-star hotels. In general, it may be said that relative customer satisfaction was higher for 4 and 5-star hotels (as expected).

Note that this ranking is based solely on the opinions of those who have submitted their comments based on 12 aspect categories. In

Table 20
Prioritizing of the 5-star hotels using BWM concerning ABSA of crawled reviews.

Hotel Name	Score	Rank	Hotel Name	Score	Rank
Meriton Suites Zetland	1.07	24	Sir Stamford at Circular Quay Hotel Sydney	1.28	18
Amora Hotel Jamison Sydney	0.83	29	Meriton Suites World Tower, Sydney	0.76	31
The Old Clare Hotel	1.59	9	Larmont Sydney by Lancemore	2.11	3
Meriton Suites Pitt Street, Sydney	0.91	27	The Darling	1.42	13
Meriton Suites Kent Street, Sydney	2.04	4	Radisson Blu Plaza Hotel Sydney	2.01	6
Sydney Harbour Marriott Hotel at Circular Quay	1.28	17	Pier One Sydney Harbour, Autograph Collection	1.49	10
Pullman Quay Grand Sydney Harbour	1.20	22	Ovolo Woolloomooloo	2.48	1
Quay West Suites Sydney	0.47	33	InterContinental Sydney Double Bay	0.75	32
Harbour Rocks Hotel Sydney - MGallery Collection	1.86	7	Swissotel Sydney	2.01	5
Hilton Sydney	1.20	23	Sheraton Grand Sydney Hyde Park	1.25	20
The Langham, Sydney	1.46	11	Fraser Suites Sydney	0.47	35
Meriton Suites Waterloo	0.95	26	Primus Hotel Sydney	0.47	36
The Star Grand Hotel and Residences Sydney (Formerly Astral Tower and Residences)	1.36	15	InterContinental Sydney	0.47	37
Sofitel Sydney Wentworth	0.97	25	Sofitel Sydney Darling Harbour	0.47	38
Meriton Suites Sussex Street Sydney	0.47	34	Four Seasons Hotel Sydney	0.81	30
Meriton Serviced Apartments George Street	2.23	2	Park Hyatt Sydney	1.27	19
Pullman Sydney Hyde Park	0.85	28	QT Bondi	1.46	12
Establishment Hotel	1.20	21	Meriton Suites Campbell Street Sydney	1.42	14
Hyatt Regency Sydney	1.35	16	Shangri-La Hotel Sydney	0.47	39
QT Sydney	1.61	8			

practice, by varying the preference values of each aspect over the others, the hotel ranking will be different in each case study, and the final ranking will be directly dependent on the customer's opinion and taste.

6. Theoretical and practical implications

Semantic mining and sentiment analysis are some of the fields that have been researched in recent years. ABSA is a category of sentiment analysis for extracting opinion terms and aspect terms and the relations between them, mainly from web-based platforms. Polarity Detection has often been considered by researchers. While ignoring the effect of Aspect term extraction, Aspect category detection on Polarity Detection reduces the validity of the results. Unlike previous studies that examined only one step, this study considers all three steps simultaneously. On the other hand, for the first and second steps, the rule-based approach has been used to achieve the highest accuracy. Using this approach for these two steps increases the accuracy in the third step significantly. It also allows the decision-making system designed in the final step to provide results that are close to reality. One of the gaps in the ABSA research is the lack of decision-making methods. So that ABSA outputs can be used in the decisions of different groups such as policymakers, tourism service companies and tourists. By integrating decision-making methods with ABSA, a strong decision support system can be achieved. In this research, ABSA output enters into the BWM method without user intervention and will lead to the ranking of hotels at different levels. In other words, the complete end-to-end ranking system integrating ABSA and MCDM techniques is provided.

Considering the practical aspect is another highlight of the research that can be mentioned. With the spread and pervasiveness of the Internet and the large volume of day-to-day business such as sales, communications and office work conducted on online platforms, vast amounts of data are exchanged in this space. Users' feedback on the products or services received are valuable sources of real and unmediated feedback and contain important information. This information is important for both the recipient and the provider of the product or service. From the perspective of the providers of goods or services and due to the fierce competition resulting from globalization, the processing and analysis of this information can improve customer satisfaction with the services or

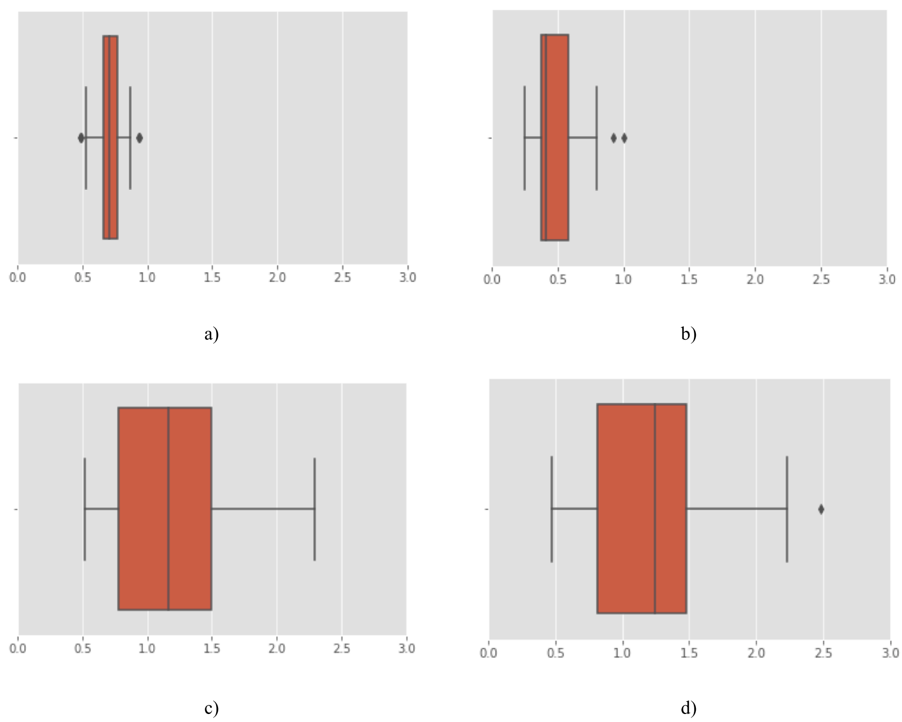

Fig. 4. Box plot of calculated scores: a) 2-star hotels, b) 3-star hotels, c) 4-star hotels, d) 5-star hotels.

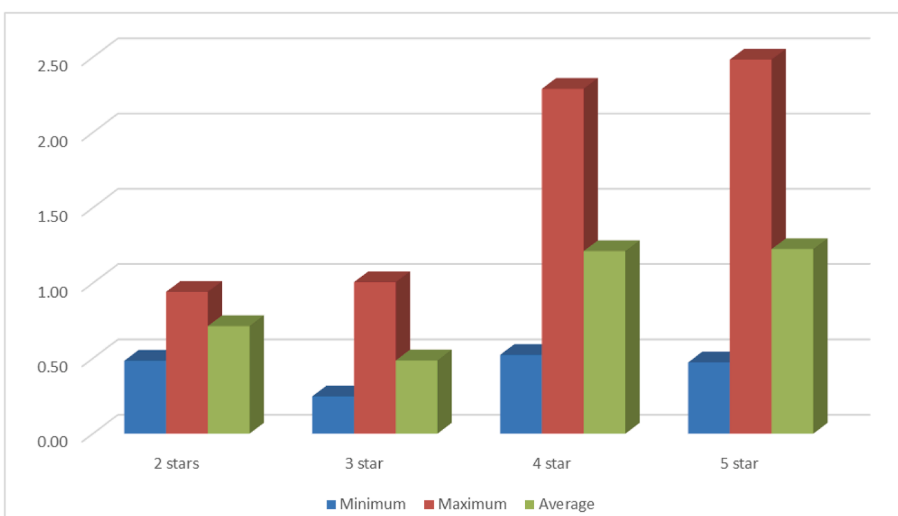

Fig. 5. Comparison of calculated scores for different hotel groups.

Table 21
Wilcoxon signed-rank test.

	Minimum	Maximum	Average
Difference	0.125	0.125	0.250

products received and create a good competitive advantage for companies. From the perspective of the customers or the recipients of goods or services, the recorded opinions of other people's experiences will be a useful guide to making the right choice. However, due to the high volume of data and shortcomings such as mental fatigue due to the effort required to review the users' comments, the time-consuming review process, the inability of the brain to process large amounts of comments simultaneously and the reader's unfamiliarity with the concepts and terms used in the comments, it is almost impossible to manually process this data. For this reason, there is a critical need for automated end-to-end processing systems in the analysis of this data. One of the most important types of end-to-end processing is end-to-end ranking. In the end-to-end ranking, by combining mechanisms such as text processing, sentiment analysis, dimensionality reduction, and MCDM techniques, a comprehensive approach is followed to convert unstructured and incoherent big data into a user-friendly ranking. The main advantage of end-to-end ranking over intermediate rankings (experts or intermediary companies) is due to the realistic views of customers. In this approach, the basic data is obtained directly from the opinions of users and customers, and without the intervention of personal opinions, the final information is prepared using the criteria and preferences of end-users or customers.

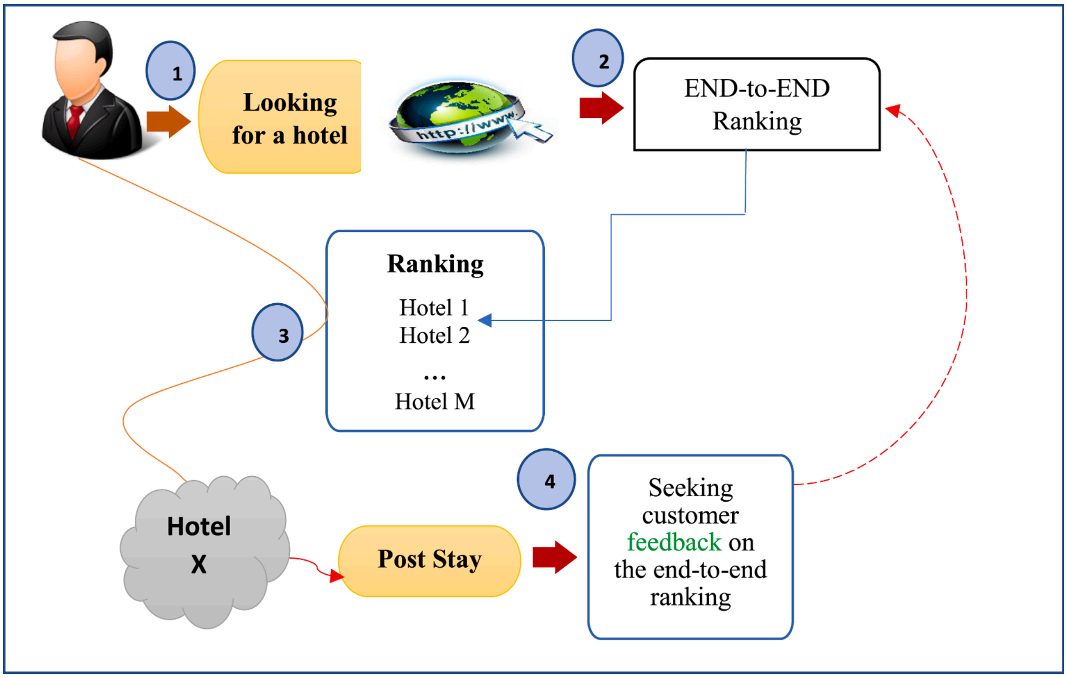

Fig. 6. User-centered-based evaluation.

The final result of the ranking will be highly dependent on the preferences of customers, and as the preferences of each customer change, the ranking output will change. The main advantage of the approach is that according to the taste of each customer and relying on the opinions of other customers it provides a logical ranking based on taste and reality. As a result, the output ranking result of the approach cannot be compared to that of the TripAdvisor online platform.

7. User-centered based evaluation: An agenda for future research

The AI-powered product designer should consider the product users and their feedback on the architecture of the product to make sure it will meet the needs of the users. In this work, we proposed a new end-to-end ranking model which ranks accommodation based on the customers' priorities and tastes, relying on external information. The model provides a service for tourists and is the base of the AI-powered hotel ranking model. While traditional AI models have been verified through various metrics by dividing the available data into training and testing datasets, due to the nature of live stream data which varies greatly in terms of quality and patterns, and the variation in hotel services due to internal and external circumstances, evaluating AI models using traditional methods is no longer sufficient. The evaluation should go beyond the optimization of loss functions. One way to do this is to build this model considering the user.

Unfortunately, we did not have the opportunity to make the proposed end-to-end ranking model a product, thereby its performance has been evaluated using statistical-based metrics. However, to determine how these types of models should be evaluated in practice, we outline a user-centered-based evaluation method. As depicted in Fig. 6, the feedback of the customer after using the ranking model should be obtained and thereby the performance of the ranking method is evaluated from this perspective.

8. Conclusion

Choosing a service or product is one of the challenges that customers face. The choice of product or service according to the rankings made by third parties in the current systems usually does not satisfy customers. Particularly about the selection of services, such as a tourist's choice of a hotel, different tastes and opinions can be considered by customers. This article provides an end-to-end ranking framework in which the opinions of customers who have stayed in a particular hotel are taken into account. For this purpose, customers' reviews are crawled on the relevant sites and after pre-processing, the related aspect terms and aspect categories using SemEval-2016 are extracted using the defined rule-based method, and then the polarity of the reviews is determined using machine learning methods. The users' opinions were aggregated for each aspect category per each hotel in each star category using the Dawid-Skene algorithm. These polarities are then used in matrix form as input to the BWM method and then the hotels are ranked. One of the strengths of this method is that it takes into account all the opinions of previous customers but it does not take into account the opinions of third parties who may have certain orientations or interests from which they will receive a benefit. This approach also provides an automated framework for customers (tourists) so that each customer can make this ranking based on their preferences for certain aspects. This can be done in different time frames. The proposed end-to-end ranking framework is based on customer reviews of Sydney hotels and explains the analytics and benefits of the proposed approach. In future endeavours, the effects of gender and the nationality of the tourists who have used these hotels can be considered. Also, the effect of special dates such as holidays, busy seasons, etc. can be considered in this ranking.

CRediT authorship contribution statement

Milad Eshkevari: Methodology, Software, Formal analysis, Visualization, Writing – original draft. **Mustafa Jahangoshai Rezaee:** Conceptualization, Methodology, Writing – original draft, Supervision. **Morteza Saberi:** Conceptualization, Investigation, Validation. **Omar Hussain:** Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

References

Aboutorab, H., Saberi, M., Asadabadi, M. R., Hussain, O., & Chang, E. (2018). ZBWM: The Z-number extension of Best Worst Method and its application for supplier development. *Expert Systems with Applications, 107*, 115–125.

Agarwal, A., Mandal, D., Parkes, D. C., & Shah, N. (2020). Peer prediction with heterogeneous users. *ACM Transactions on Economics and Computation (TEAC), 8*(1), 1–34.

Alamoudi, E. S., & Alghamdi, N. S. (2021). Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings. *Journal of Decision Systems*, 1–23.

Ali, T., Marc, B., Omar, B., Soulaimane, K., & Larbi, S. (2021). Exploring destination's negative e-reputation using aspect based sentiment analysis approach: Case of Marrakech destination on TripAdvisor. *Tourism Management Perspectives, 40*, Article 100892.

Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., & Gupta, B. (2018). Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *Journal of Computational Science, 27*, 386–393.

Al-Smadi, M., Talafha, B., Al-Ayyoub, M., & Jararweh, Y. (2019). Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *International Journal of Machine Learning and Cybernetics, 10*(8), 2163–2175.

Arenas-Márquez, F. J., Martinez-Torres, R., & Toral, S. (2021). Convolutional neural encoding of online reviews for the identification of travel group type topics on TripAdvisor. *Information Processing & Management, 58*(5), Article 102645.

Chen, X., Sun, C., Wang, J., Li, S., Si, L., Zhang, M., & Zhou, G. (2020). Aspect sentiment classification with document-level sentiment preference modeling. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Chiu, D. K., Cheung, S. C., Hung, P. C., Chiu, S. Y., & Chung, A. K. (2005). Developing e-negotiation support with a meta-modeling approach in a web services environment. *Decision Support Systems, 40*, 51–69.

Dawid, A. P., & Skene, A. M. (1979). Maximum likelihood estimation of observer error-rates using the EM algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 28*, 20–28.

Devlin, J., Chang, M. -W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 1810.04805.

Feichtenhofer, C., Pinz, A., & Zisserman, A. (2017). Detect to track and track to detect. *In Proceedings of the IEEE International Conference on Computer Vision (3038–3046)*.

García-Pablos, A., Cuadros, M., & Rigau, G. (2018). W2VLDA: Almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications, 91*, 127–137.

Huang, J., Meng, Y., Guo, F., Ji, H., & Han, J. (2020). Weakly-supervised aspect-based sentiment analysis via joint aspect-sentiment topic embedding. arXiv preprint arXiv: 2010.06705.

Huang, Y., Li, C., Wu, J., & Lin, Z. (2018). Online customer reviews and consumer evaluation: The role of review font. *Information & Management, 55*(4), 430–440.

Iqbal, M., Subedi, N., & Aryafar, K. (2019). Production Ranking Systems: A. *Review.* arXiv preprint arXiv:1907.12372.

Karimi, A., Rossi, L., & Prati, A. (2020). Improving BERT Performance for Aspect-Based Sentiment Analysis. arXiv preprint arXiv:2010.11731.

Kim, J., & Lim, C. (2021). Customer complaints monitoring with customer review data analytics: An integrated method of sentiment and statistical process control analyses. *Advanced Engineering Informatics, 49*, Article 101304.

Kumar, A., Verma, S., & Sharan, A. (2021). ATE-SPD: Simultaneous extraction of aspect-term and aspect sentiment polarity using Bi-LSTM-CRF neural network. *Journal of Experimental & Theoretical Artificial Intelligence, 33*(3), 487–508.

Laddha, A., & Mukherjee, A. (2018). Aspect opinion expression and rating prediction via LDA–CRF hybrid. *Natural Language Engineering, 24*(4), 611–639.

Li, X., Bing, L., Zhang, W., & Lam, W. (2019). Exploiting BERT for end-to-end aspect-based sentiment analysis. arXiv preprint arXiv:1910.00883.

Li, Y., Yin, C., Wei, T., Zhong, H., Luo, J., Xu, S., & Wu, X. (2019). A joint model for aspect-category sentiment analysis with contextualized aspect embedding. arXiv preprint arXiv:1908.11017.

Li, Z., Li, L., Zhou, A., & Lu, H. (2021). JTSG: A joint term-sentiment generator for aspect-based sentiment analysis. *Neurocomputing, 459*, 1–9.

Liu, X., Liang, D., Yan, S., Chen, D., Qiao, Y., & Yan, J. (2018). Fots: Fast oriented text spotting with a unified network. *In Proceedings of the IEEE conference on computer vision and pattern recognition (5676–5685)*.

Liu, N., & Shen, B. (2020). Aspect-based sentiment analysis with gated alternate neural network. *Knowledge-Based Systems, 188*, Article 105010.

Movahedi, S., Ghadery, E., Faili, H., & Shakery, A. (2019). Aspect category detection via topic-attention network. arXiv preprint arXiv:1901.01183.

Papadimitriou, G., Wang, C., Vahi, K., da Silva, R. F., Mandal, A., Liu, Z., … Kettimuthu, R. (2021). End-to-end online performance data capture and analysis for scientific workflows. *Future Generation Computer Systems, 117*, 387–400.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A.-S., De Clercq, O. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. Paper presented at the Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016).

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega, 53*, 49–57.

Schouten, K., & Frasincar, F. (2015). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering, 28*(3), 813–830.

Shirouyehzad, H., Tavakoli, M. M., & Badakhshian, M. (2016). The Linear Assignment Method for Ranking of Organizations with Service Quality Approach: A Case Study of Hotels in City of Isfahan. Journal of Applied Research on. *Industrial Engineering, 3* (1(4)), 49–57.

Singh, K. K., & Lee, Y. J. (2016). In *End-to-end localization and ranking for relative attributes* (pp. 753–769). Cham: Springer.

Tang, H., Ji, D., Li, C., & Zhou, Q. (2020). Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. *Paper presented at the Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Tran, T., Ba, H., & Huynh, V. N. (2019, March). Measuring hotel review sentiment: An aspect-based sentiment analysis approach. In International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making (pp. 393-405). Springer, Cham.

Trușcă, M. M., Wassenberg, D., Frasincar, F., & Dekker, R. (2020). A Hybrid Approach for Aspect-Based Sentiment Analysis Using Deep Contextual Word Embeddings and Hierarchical Attention. *Paper presented at the International Conference on Web Engineering*.

Valipour, M., Yousefi, S., Rezaee, M. J., & Saberi, M. (2021). A clustering-based approach for prioritizing health, safety and environment risks integrating fuzzy C-means and hybrid decision-making methods. *Stochastic Environmental Research and Risk Assessment*, 1–20.

Wallaart, O., & Frasincar, F. (2019). A Hybrid Approach for Aspect-Based Sentiment Analysis Using a Lexicalized Domain Ontology and Attentional Neural Models. *Paper presented at the European Semantic Web Conference*.

Wang, X., Leng, M., Song, J., Luo, C., & Hui, S. (2019). Managing a supply chain under the impact of customer reviews: A two-period game analysis. *European Journal of Operational Research, 277*(2), 454–468.

Wang, X., Tang, M., Yang, T., & Wang, Z. (2021). A novel network with multiple attention mechanisms for aspect-level sentiment analysis. *Knowledge-Based Systems, 227*, Article 107196.

Wu, W., Liu, G., Ye, H., Zhang, C., Wu, T., Xiao, D., Lin, W. & Zhu, X. (2018). EENMF: An End-to-End Neural Matching Framework for E-Commerce Sponsored Search. arXiv preprint arXiv:1812.01190.

Xu, K., Liao, S. S., Li, J., & Song, Y. (2011). Mining comparative opinions from customer reviews for competitive intelligence. *Decision Support Systems, 50*(4), 743–754.

Xu, X., Lu, H., Song, J., Yang, Y., Shen, H. T., & Li, X. (2019). Ternary adversarial networks with self-supervision for zero-shot cross-modal retrieval. *IEEE transactions on cybernetics, 50*(6), 2400–2413.

Xu, X. (2021). Closed-form evaluations and open-ended comment options: How do they affect customer online review behavior and reflect satisfaction with hotels? *Decision Support Systems, 145*, Article 113525.

Yu, H., Huang, Y., Pi, L., Zhang, C., Li, X., & Wang, L. (2021). End-to-end video text detection with online tracking. *Pattern Recognition, 113*, Article 107791.

Yu, T., Zhang, Y., & Lin, K. J. (2007). Efficient algorithms for web services selection with end-to-end qos constraints. *ACM Transactions on the Web (TWEB), 1*(ISO 9001), 6-es.

Zhang, C., Li, Q., & Song, D. (2019). Aspect-based sentiment classification with aspect-specific graph convolutional networks. arXiv preprint arXiv:1909.03477.

Zhang, M., Fan, B., Zhang, N., Wang, W., & Fan, W. (2021). Mining product innovation ideas from online reviews. *Information Processing & Management, 58*(1), Article 102389.

Zhang, Y., Xu, B., & Zhao, T. (2020). Convolutional multi-head self-attention on memory for aspect sentiment classification. *IEEE/CAA Journal of Automatica Sinica, 7*(4), 1038–1044.