

Received December 6, 2021, accepted December 28, 2021, date of publication December 31, 2021, date of current version January 7, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3139956

# Cross-Platform Reputation Generation System Based on Aspect-Based Sentiment Analysis

**ACHRAF BOUMHIDI<sup>1</sup>**, **ABDESSAMAD BENLAHBIB<sup>1</sup>**, AND **EL HABIB NFAOUI<sup>1</sup>**, (Member, IEEE)

LISAC Laboratory, Computer Science Department, Faculty of Sciences Dhar EL Mehraz (F.S.D.M), Sidi Mohamed Ben Abdellah University, Fes 30003, Morocco

Corresponding author: Achraf Boumhidi (achraf.boumhidi@usmba.ac.ma)

**ABSTRACT** The active growth of Internet-based applications such as social networks and e-commerce websites leads people to generate a tremendous amount of opinions and reviews about products and services. Thus, it becomes very crucial to automatically process them. Over the last ten years, many systems have been proposed to generate and visualize reputation by mining textual and numerical reviews. However, they have neglected the fact that online reviews could be posted by malicious users that intend to affect the reputation of the target product. Besides, these systems provide an overall reputation value toward the entity and disregard generating reputation scores toward each aspect of the product. Therefore, we developed a system that incorporates spam filtering, review popularity, review posting time, and aspect-based sentiment analysis to generate accurate and reliable reputation values. The proposed model computes numerical reputation values for an entity and its aspects based on opinions collected from various platforms. Our proposed system also offers an advanced visualization tool that displays detailed information about its output. Experiment results conducted on multiple datasets collected from various platforms (Twitter, Facebook, Amazon . . . ) show the efficacy of the proposed system compared with state-of-the-art reputation generation systems.

**INDEX TERMS** Aspect-based sentiment analysis, decision-making, reputation generation, e-commerce, opinion mining.

## I. INTRODUCTION

Having easy access to the web has radically changed the way people interact with brands and products. From physical products to online services, people tend to instantly share their opinions and reviews on various platforms on the Internet. A recent research experiment<sup>1</sup> shows that consumers are more willing to share a review when the experience they have had evokes emotions, whether positive or negative. This large volume of consumers' reviews holds insightful information about the quality of the product/service, therefore analyzing them will help consumers make a better judgment toward the targeted item. In the past few years, a new sub-field of natural language processing (NLP) called reputation generation has been well-established as an area of interest. The main focus of reputation generation systems is to produce a numerical value in which an entity is held based on mining customer reviews and their numerical ratings.

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine<sup>1</sup>.

<sup>1</sup><https://business.trustpilot.com/reviews>

Over the last decade, many reputation generation systems have been proposed [1]–[8] to generate and visualize reputation of online products and services based on fusing and mining textual and numerical reviews. However, these systems have not taken into consideration (1) extracting and processing reviews from various platforms, (2) filtering reviews written by potential spammers, (3) generating a numerical reputation value toward each aspect of the target product, and, (4) providing an advanced reputation visualization tool for a better decision-making process. Thereby, we designed and built an upgraded reputation generation model that overcomes the shortcomings of the previous systems in order to compute and visualize the reputation of an entity (product, movie, hotel, restaurant, service) with consistent reliability. The proposed system collects and processes data from both e-commerce and social media platforms. Then, a spam filtering system is applied to eliminate spam reviews and prepare the cleaned output for aspect-based sentiment analysis (ABSA), where aspects of the target entity are extracted from the reviews with their sentiment polarities. Later, the time and popularity features of the reviews are exploited along with

the ASBA results to finally generate a reputation value of each aspect of the target entity as well as the overall reputation value using mathematical formulas. The system also proposes an analytical dashboard that displays in-depth information about the reputation of the target entity.

In this manner, this study addresses the following research question: with the consideration of review popularity, review time, spam filtering, and ABSA, can the proposed reputation model offer better results in terms of generating and visualizing reputation than state-of-the-art (SOTA) systems?

This paper is organized as follows. Section 2 presents the related work concerning the previous reputation generation systems as well as the ABSA models. Section 3 presents the preliminaries. Section 4 describes our proposal. Section 5 details the experiments. Section 6 presents the discussion. And finally, Section 7 concludes this paper.

## II. RELATED WORK

This section reviews the work done in the field of ABSA, and reputation generation systems based on NLP techniques.

### A. ASPECT-BASED SENTIMENT ANALYSIS

Sentiment Analysis (SA), also defined as opinion mining is one of the rapidly growing research areas in the past few years [9] that aims to extract the polarity of an entity. SA can generally occur at different levels: Document-level [10], sentence-level [11], and aspect-level [12]. This sub-section will focus on ABSA since it is applied in this paper. ABSA identifies the aspects in the given textual review about a product/service and determines which class of sentiment those aspects belong to. ABSA can be categorized by two main processing phases: Aspect extraction (AE), and Aspect polarity classification (APC). The first phase deals with the extraction of aspects, which can be aspect terms [13], explicit aspects [14], and implicit aspects [15]. The second phase sentimentally classifies the predefined aspects into positive, negative, or neutral. In [16], the authors were the first to propose a set of techniques for mining and summarizing product reviews based on NLP. Their main objective was to provide a feature-based summary of a large number of customer reviews of an online product. They started by mining product features that have been expressed by the customers using the association rule mining algorithm [17]. Next, they identified the opinion sentences in each review in order to determine the polarity of each opinion sentence. Finally, they produced a summary using the discovered information. Further, in [18], Poria *et al.* presented the first deep learning approach for the AE task in opinion mining. The authors employed a 7-layer deep convolutional neural network to tag each word in the textual opinions as either aspect or non-aspect word. The authors also proposed a set of heuristic linguistic patterns and integrated them with the deep learning classifier which significantly improves the accuracy compared with the previous SOTA methods. In [19], the authors proposed an attention-based long short-term memory (LSTM) [20] for aspect-level sentiment classification. The idea is to learn

aspect embeddings and let aspects participate in computing attention weights. The proposed model can focus on different parts of a sentence when different aspects are given so that they are more competitive for aspect-level classification. The proposed model achieved better results compared with the standard LSTM on the SemEval 2014 Task 4 dataset [21]. In [22], Wei and Toi improved the deficiencies of the previous LSTM approaches by proposing convolutional neural networks [23] and gating mechanisms (GCAE) based model, which has been proved to be more accurate and efficient. The novel Gated Tanh-ReLU Units can selectively output the sentiment features according to the provided aspect or entity. The architecture of the proposed model is much simpler than the attention layer used in the previously existing models. The experiments on SemEval datasets show a performance improvement compared with the LSTM based models. The authors in [24] proposed an interactive multi-task learning network (IMN) capable of jointly learning multiple related tasks simultaneously at both the token-level as well as the document-level. The IMN introduces a message passing mechanism that allows informative interactions between tasks, enabling the correlation to be better exploited. Experiments on three benchmark datasets, taken from SemEval2014 and SemEval 2015 [25] show that IMN outperforms other baselines by large margins. Since most existing methods ignore the position information of the aspect when encoding the sentence, authors in [26] proposed a hierarchical attention-based position-aware network (HAPN), which includes position embeddings to learn the position-aware representations of sentences to generate the target-specific representations of contextual words. HAPN achieved the SOTA performance on SemEval 2014 dataset compared with the previous methods. Xu *et al.* [27] presented a review reading comprehension (RRC) task where they adopted BERT [28] as a base model, and proposed a joint post-training and fine-tuning approach for ATE, APC. Experimental results show that the proposed post-training approach is very effective. Later in [29], the authors proposed a novel architecture called BERT Adversarial Training (BAT) to employ adversarial training for AE and APC by generating artificial data which is carried out in the embedding space. The proposed model outperforms the standard BERT as well as the in-domain post-trained BERT in both AE and APC tasks. In [30], the authors exploit domain-specific BERT language model finetuning in addition to supervised task-specific finetuning to produce a new SOTA performance on the SemEval 2014 Task 4 restaurants dataset. The authors also showed that cross-domain adapted BERT model performs better than strong baseline models such as XLNet-base [31] and vanilla BERT-base. In [32], the authors compared the induced trees from pre-trained models and the dependency parsing trees on various popular models for the ABSA task. They found that the induced tree from finetuned RoBERTa [33] (FT-RoBERTa) outperforms the parser-provided tree. The experiments show that the RoBERTa-based model can outperform or approximate the previous SOTA performances

on six datasets across four languages including SemEval 2014 task 4. Recently, authors in [34] proposed a multi-task learning model named LCF-ATEPC for ABSA based on the multi-head self-attention and the local context focus (LCF) [35] mechanisms. The proposed model is multi-lingual and applicable to the classic English review SA task, such as the SemEval-2014 task4. The proposed model can automatically extract aspects and determine their sentiment polarities. Since the LCF-ATEPC model currently achieves SOTA performance on AE and APC tasks,<sup>2</sup> it was selected to be employed in this paper.

## B. REPUTATION GENERATION

The Oxford Learner's Dictionaries<sup>3</sup> defines reputation as “the opinion that people have about what someone or something is like, based on what has happened in the past”. In the 21st century, several reputation systems have been proposed to compute a satisfaction score toward various online items [36]–[42] including movies, TV shows, hotels, and products. Surprisingly enough, these systems have relied only on numerical reviews (ratings) during reputation computation and have disregarded the exploitation of textual reviews until 2012 when Abdel-Hafez et al. [1] designed a reputation model that uses reviews expressed in natural languages rather than users' ratings to compute a realistic reputation value for every feature of the product and the product itself by incorporating opinion orientation and opinion strength (opinion mining), nevertheless, no evidence has been provided to support the efficiency of their product reputation system. Yan et al. [3] proposed the first system that combines opinion fusion and semantic analysis to generate and visualize reputation toward Amazon's products. This system has been lately improved in [4] by adding a binary sentiment classification step before the opinion fusion and grouping phase. Benlahbib and Nfaoui [6] designed and built a reputation model that considers review time, review helpfulness, and review sentiment intensity during reputation computation and visualization. Elmurungi and Gherbi [5] proposed a system that computes reputation scores from users' feedback based on a SA model. The reputation score of a product is the ratio of the number of positive reviews over the total number of reviews toward this product. The same idea was adopted in [43], [44]. In [45], the authors presented a reputation system dedicated to movies and TV shows. The model integrates fine-grained opinion mining (Multinomial Naïve Bayes classifier trained on the SST-5 dataset [46]) and semantic analysis (Embeddings from Language Models (ELMo) [47]) to generate a realistic reputation value from user's reviews. Recently, Boumhidi and Nfaoui [8] proposed the first system that generates a reputation value toward various entities (movies, products, hotels, and restaurants) from user-generated data posted on Twitter microblogging

<sup>2</sup><https://paperswithcode.com/sota/aspect-based-sentiment-analysis-on-semeval>

<sup>3</sup>[https://www.oxfordlearnersdictionaries.com/definition/american\\_english/reputation](https://www.oxfordlearnersdictionaries.com/definition/american_english/reputation)

website. The system applied a Bidirectional Encoder Representations from Transformers (BERT) classifier to extract the sentiment orientation of the textual tweets. Next, a sentiment intensity score is calculated from the positive tweets. Finally, the proposed system incorporated the previous results with a computed popularity score from the extracted Twitter features (number of followers, account authenticity, number of likes, number of retweets) to generate a single numerical reputation value between 0 and 10.

TABLE 1 summarizes the opinion mining techniques exploited during the reputation generation and visualization process for the previous reputation systems and for ours.

## III. PROBLEM STATEMENT

The purpose of this research is to compute a numerical reputation value for a specific entity as well as the reputation of each aspect of that entity. Let's assume that the target entity is a phone product, the goal is to generate the overall reputation value for the phone e.g., “7/10” based on textual reviews collected from various platforms, as well as generating a reputation value for the aspects of the phone e.g., “camera: 5/10”, “design: 9/10”, etc. The collected set of reviews  $R'_j = \{r_{1jk_1}, r_{2jk_2}, \dots, r_{mjk_p}\}$  for an entity  $E_j$  posted by a set of users  $U_j = \{u_{jk_1}, u_{jk_2}, \dots, u_{jk_p}\}$  is passed through the process of review spam filtering where the output results is a free-spam set of reviews  $R_j = \{r_{1jk_1}, r_{2jk_2}, \dots, r_{njk_o}\}$  ( $n \leq m$ ), that will be passed to an ABSA model named LCF-ATEPC for the purpose of extracting the entity aspects and their sentiment orientations for each review in the  $R_j$  set. Similar aspects are then grouped jointly with their polarities, and by using mathematical formulas, the previous results are incorporated with a calculated set of review time scores  $TSR_j = \{tsr_{1j}, tsr_{2j}, \dots, tsr_{nj}\}$  and a set of review popularity scores  $PSR_j = \{psr_{1j}, psr_{2j}, \dots, psr_{nj}\}$  to finally generate reliable and trustworthy reputation values. The set of review popularity scores  $PSR_j$  is computed based on the set of review likes  $L_j = \{l_{1j}, l_{2j}, \dots, l_{nj}\}$  and the set of review shares  $S_j = \{s_{1j}, s_{2j}, \dots, s_{nj}\}$ .

## IV. PROPOSED APPROACH

This section is divided into eight subsections that describe the architectural overview of the proposed system, data collection and processing, opinion spam detection, aspect extraction and classification, popularity score calculation, time score calculation, reputation generation, and finally, reputation visualization, respectively.

### A. SYSTEM OVERVIEW

This system aims at generating a reputation value toward online entities (movies, hotels, restaurants, services, etc.) and computing a satisfaction score toward each aspect of the target entity by processing textual and numerical data collected from multiple platforms. FIGURE 1 presents its architecture. First, we start by collecting users' reviews from different platforms such as Twitter, Amazon, YouTube, etc. Next, an automatic spammers filtering system is employed to

**TABLE 1.** Summary of reputation systems based on NLP techniques.

Work	Language	Domain	Semantic Analysis	Document-Level Sentiment Analysis	Aspect-Based Sentiment Analysis
Abdel-Hafez et al. (2012) [1]	N/A	N/A	N/A	N/A	N/A
Farooq et al. (2016) [2]	English	Products	N/A	N/A	<ul style="list-style-type: none"> <li>• Association rule mining</li> <li>• SentiWordNet [48], [49]</li> </ul>
Yan et al. (2017) [3]	<ul style="list-style-type: none"> <li>• English</li> <li>• Chinese</li> </ul>	Products	Latent Semantic Analysis (LSA)	N/A	N/A
Benlahbib and Nfaoui (2019) [50]	English	Movies	Latent Semantic Analysis (LSA)	N/A	N/A
Benlahbib et al. (2019) [51]	English	Movies	Latent Semantic Analysis (LSA)	Logistic Regression	N/A
Benlahbib and Nfaoui (2020) [4]	English	Movies	Latent Semantic Analysis (LSA)	<ul style="list-style-type: none"> <li>• Naïve Bayes</li> <li>• Linear Support Vector Machine</li> </ul>	N/A
Elmurngi and Gherbi (2020) [5]	English	Products	N/A	Logistic Regression	N/A
Benlahbib and Nfaoui (2020) [6]	English	<ul style="list-style-type: none"> <li>• Products</li> <li>• Movies &amp; TV Shows</li> <li>• Hotels</li> </ul>	N/A	Bidirectional Encoder Representations from Transformers (BERT)	N/A
Benlahbib and Nfaoui (2020) [52]	English	Products	N/A	Bidirectional Gated Recurrent Unit (Bi-GRU)	N/A
Boumhidi and Nfaoui (2020) [43]	English	<ul style="list-style-type: none"> <li>• Movies</li> <li>• Restaurants</li> </ul>	N/A	Bidirectional Gated Recurrent Unit (Bi-GRU)	N/A
Gupta et al. (2020) [7]	English	<ul style="list-style-type: none"> <li>• Movies</li> <li>• Books</li> </ul>	N/A	<ul style="list-style-type: none"> <li>• Bidirectional Encoder Representations from Transformers (BERT)</li> <li>• Naïve Bayes</li> <li>• Support Vector Machine</li> </ul>	N/A
Boumhidi et al. (2021) [44]	English	Movies	N/A	Bidirectional Long Short-Term Memory (Bi-LSTM)	N/A
Benlahbib and Nfaoui (2021) [45]	English	Movies & TV Shows	Embeddings from Language Models (ELMo)	Multinomial Naïve Bayes	N/A
Boumhidi and Nfaoui (2021) [8]	English	<ul style="list-style-type: none"> <li>• Products</li> <li>• Services</li> <li>• Hotels</li> <li>• Movies</li> </ul>	N/A	Bidirectional Encoder Representations from Transformers (BERT)	N/A
This study	English	<ul style="list-style-type: none"> <li>• Products</li> <li>• Services</li> <li>• Hotels</li> <li>• Movies</li> </ul>	N/A	N/A	LCF-ATEPC

detect and eliminate unwanted spam reviews. Then, we apply a SOTA ABSA model to users' textual reviews in order to compute a score based on the sentiment orientation of the extracted aspects from those reviews. Further, we calculate a popularity score and a time score based on statistical features extracted with the textual reviews. Finally, we compute a reputation value based on the previously calculated scores, and we propose a new user-friendly visualization interface that displays in-depth details about the reputation of the target entity.

### B. DATA COLLECTION AND PREPROCESSING

One of the important features of the proposed system is the ability to collect and process data from various platforms. Previous reputation generation systems gather necessary data from either e-commerce websites such as Amazon,

TripAdvisor, or social media platforms such as Twitter and Facebook. In this work, we decided to normalize the features of all platforms in order to create a single merged dataset by classifying the platforms on the Internet into two types: the first type provides the accessibility of extracting the textual review with the number of likes received for that review such as Amazon, YouTube, etc. The second type provides the accessibility of extracting the textual review with the number of likes received for that review along with the number of times the review was shared among the network such as Twitter, Facebook, etc.

By using web scraping tools, we have collected data from both types of platforms to create a merged dataset that contains: "user name", "review text", "review time", "review likes", "review shares", and "review host". We have assigned "0" to "review shares" for reviews that are

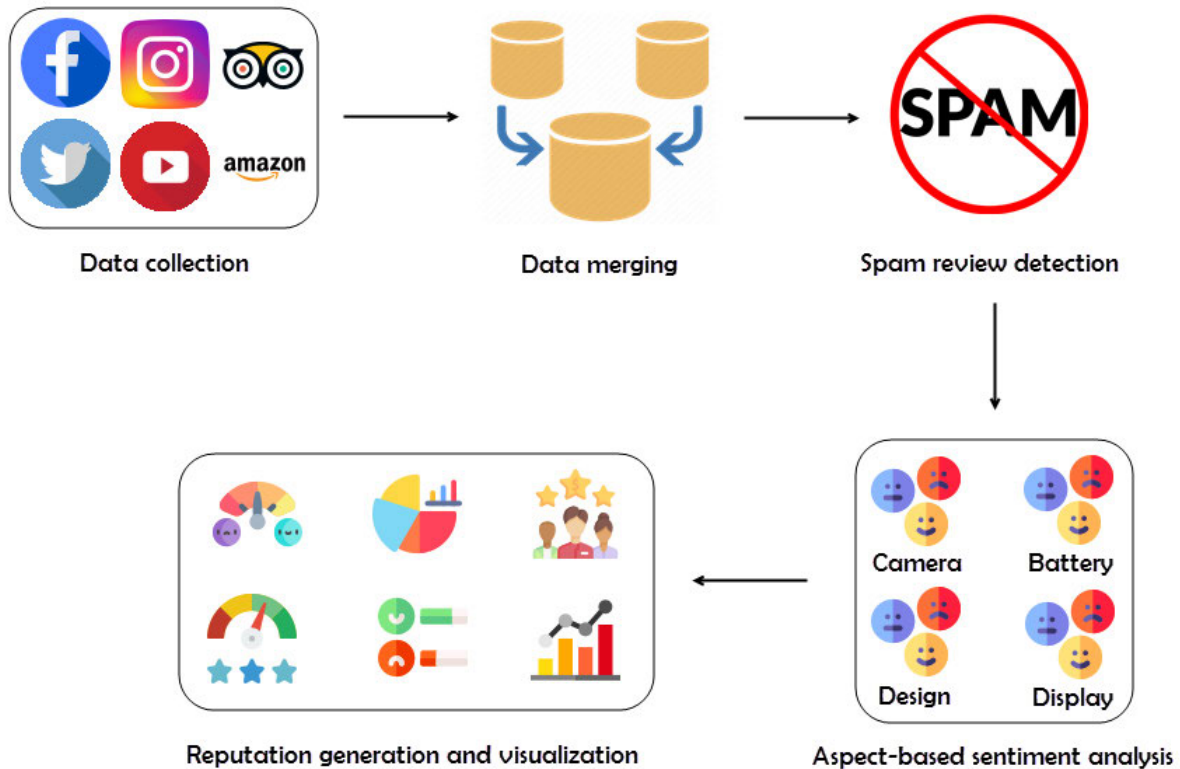


FIGURE 1. Proposed system overview.

extracted from platforms of the first type since they only provide the number of likes.

The textual reviews are cleaned using NLP techniques (text normalization, lower-casing, noise removal, etc.).

### C. OPINION SPAM DETECTION

One of the main drawbacks of opinion-sharing platforms is that anyone from anywhere in the world can post reviews about products or services without any boundaries. Opinion spammers aim to manipulate customers' opinions by either promoting or demoting the reputation of the target entity, thereby misleading the consumers [53]. Filtering and eliminating spam reviews is very critical for our reputation system in order to produce a trusted and reliable reputation value that leads to a safe decision-making process for the customers. Commercial review hosting sites e.g. Amazon and Yelp have already put through remarkable progress in detecting spam reviews [54]. However, since we are collecting peoples' opinions from multiple platforms, we choose to detect spam reviews using two normalized spammer behavioral features [55]. Notations used in this sub-section are listed in TABLE 2.

#### 1) AUTHOR CONTENT SIMILARITY (CS)

Spammers regularly post reviews that are identical or near-identical to their previous reviews, since it is a

TABLE 2. Notions used in sub-section: Opinion spam detection.

Variable	Description
$E_j$	Target entity $j$
$R'_{jk}$	Set of reviews for an entity $j$ posted by author $k$ before spam filtering
$N_j$	Total number of reviews toward entity $j$
$N_{jk}$	Total number of reviews posted by author $k$ for an entity $j$ .
$CP(R'_{jk})$	Set of pair-combination generated from the $R'_{jk}$ set

time-consuming activity for them to write a new spam review each time. Thus, we identify spammers by calculating the similarity of their reviews, where we determine a pair-combination of reviews from the set of reviews  $R'_{jk}$  posted by author  $k$  without repetition, then we converted them into vectors using pre-trained BERT model from Hugging-face,<sup>4</sup> and we apply the cosine similarity of each pair of reviews from the combination set  $CP(R'_{jk})$ . We obtained a single numerical score between 0 and 10 after we calculated the average of all the results obtained from the cosine function of each pair using Equation (1). This score is going to be used to calculate a spammer behavior score later in this sub-section.

$$F_1(u_{jk}) = \frac{\sum \text{Cosine}(CP(R'_{jk}))}{N_{jk}} \times 10 \quad (1)$$

<sup>4</sup><https://huggingface.co/>

## 2) USER NUMBER OF REVIEWS FREQUENCY (UNRF)

Posting too many reviews about the same entity is not considered normal behavior for genuine reviewers. A recent study shows that only 5-8% of spammers have a lesser ratio of posting reviews in a single day [56]. Our collected dataset may contain many reviews of the same user; therefore, we proposed Equation (2) to calculate the frequency of the number of reviews posted by user  $k$  toward the target entity  $j$ .

$$F_2(u_{jk}) = \frac{N_{jk}}{N_j} \times 10 \quad (2)$$

Based on the two spammer behavioral features suggested previously, Equation (3) is used to calculate the spammer score. Each author is assigned a label from the set  $L = \{\text{normal, spammer}\}$  by comparing the spammer score with a predetermined threshold  $\tau$  described in section 5. The label “normal” is used for regular reviewers, and “spammer” is used for spammers reviewers. Equation (4) is used to label each user.

$$\text{Score}(u_{jk}) = \frac{F_1(u_{jk}) + F_2(u_{jk})}{2} \quad (3)$$

$$L(a_i) = \begin{cases} \text{Spammer}, & \text{if } \text{Score}(u_{jk}) > \tau \\ \text{Normal}, & \text{if } \text{Score}(u_{jk}) < \tau \end{cases} \quad (4)$$

Whenever a user is labeled as a spammer, all of his reviews are going to be eliminated from the dataset. The newly cleaned spammers-free dataset is now ready for the next step of the proposed reputation system.

## D. ASPECT-BASED SENTIMENT ORIENTATION SCORE

The goal of this sub-section is to extract the aspects from the reviews and predict their sentiment polarities. For example, given a product review: “The camera on this phone is good but the design is bad,” the ABSA model needs to extract the aspects “camera” and “design”, and correctly determine their polarity. In this review, the consumers’ opinions on “camera” and “design” are positive and negative respectively. Therefore, we employed a multi-task learning model for ABSA, namely LCF-ATEPC.<sup>5</sup> This model combines APC task and aspect term extraction task (ATE), which implies that it is capable of extracting aspect term and inferring aspect term polarity. The input sequences are tokenized into separate tokens and each token is assigned two kinds of labels. The first label indicates whether the token belongs to an aspect, the second label marks the polarity of the tokens belonging to the aspect. LCF-ATEPC model integrates the pre-trained BERT model and applies self-attention and local context focus concept (LCF) to ABSA. FIGURE 2 depicts the network architecture of LCF-ATEPC. Its main components are described as follows:

- BERT-Shared Layer: LCF-ATEPC model deploys two independent BERT-Shared layers to extract local and global context features  $BERT^l$ , and  $BERT^g$  respectively. Both BERT-Shared layers are regarded as embedded

layers, and the fine-tuning process is conducted independently according to the joint loss function of multi-task learning.

- Multi-Head Self-Attention (MHSA): The multi-head attention mechanism helps the model to learn the words’ relevant information in different presentation subspaces. MHSA is based on multiple scale-dot attention that can be used to extract deep semantic features in the context. MHSA can avoid the negative influence caused by the long-distance dependence of the context when learning the features.
- Local Context Focus: Local context is a new technique that can be adapted to most fine-grained NLP tasks. The determination of local context relies on semantic-relative distance (SRD), which is used to identify how far a token is from the aspect in order to determine whether the context word belongs to the local context of a targeted aspect. Context-features Dynamic Mask (CMD) layer is employed to mask non-local context features learned by the  $BERT^l$  layer. With the CDM layer deployed, only the features of the less-semantic-relative context itself on the corresponding output position will be masked. The correlative representations between less-semantic-relative context words and aspects are reserved on corresponding output positions. In addition to (CDM) layer, another Context features Dynamic Weighted (CDW) layer is employed to focus on local context words. The goal of CDM is to drop the features of the non-local context completely. The features of a semantic-relative contextual word are retained intact while the features of less-semantic relative context will be weighted decay based on their SRD concerning a targeted aspect.
- Aspect Polarity Classifier: To perform the sentiment polarity classification, the LCF-ATEPC model combines the local context features and the global context features. Then, the aspect polarity classifier performs a head-pooling on the learned concatenated context features from the feature interactive learning layer. The Softmax function is applied to the hidden states on the corresponding position of the first token in the input sequence, in order to predict the sentiment polarity of the aspect.
- Aspect Term Extractor: Aspect term extractor performs the basic token-level classification for each token, which means that each token will be given a label, and a classification is performed to predict the aspects in the sentence.

Authors in [34] trained LCF-ATEPC model on commonly used ABSA datasets, including the Laptop and Restaurant datasets of SemEval-2014 Task4, and ACL Twitter social dataset. However, they trained the model on those datasets separately. In this paper, we trained the model on a mixed dataset of the three previously mentioned datasets in order to allow our system to treat reviews of different domains. LCF-ATEPC model achieved SOTA performance on the

<sup>5</sup><https://github.com/yangheng95/LCF-ATEPC>

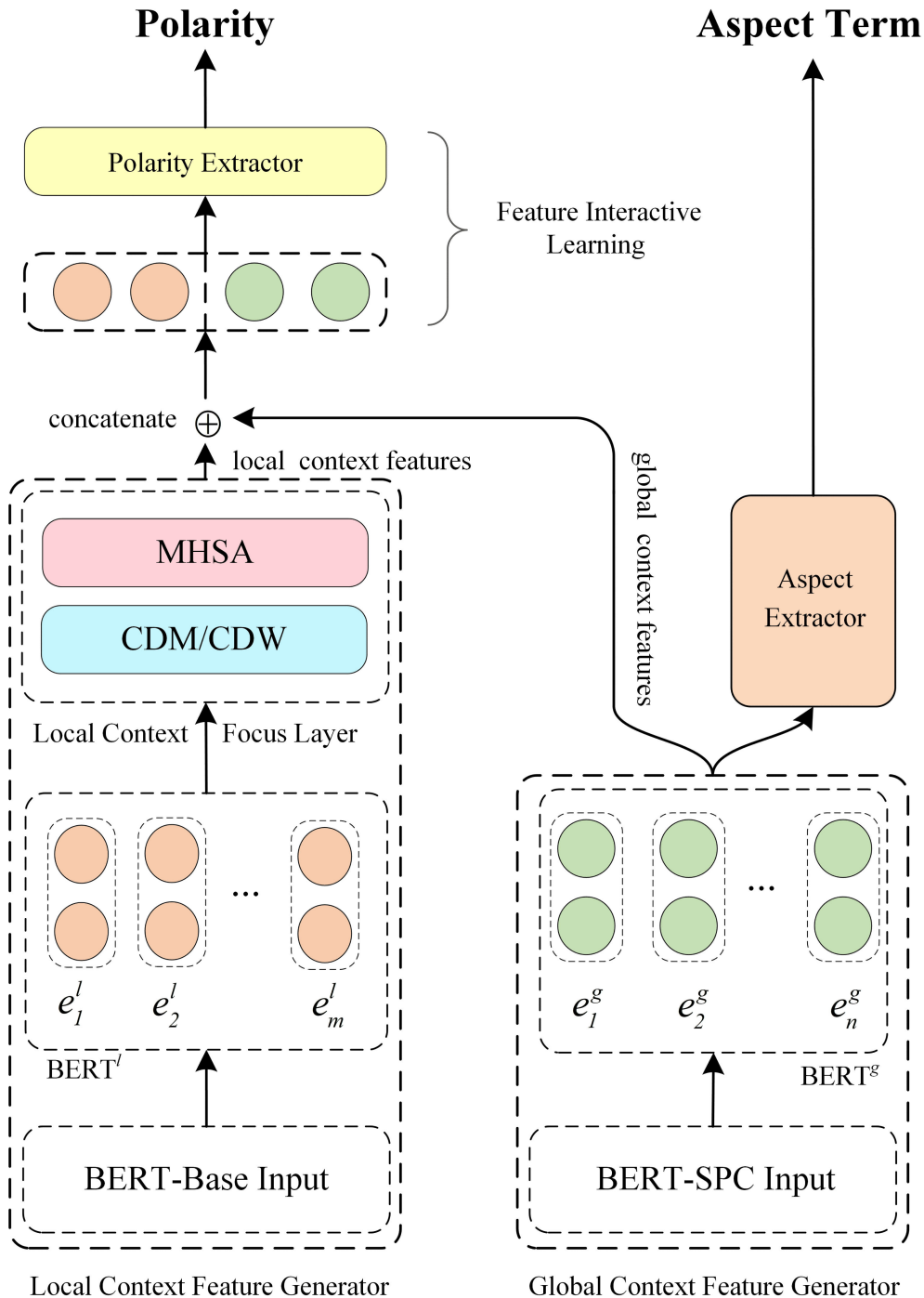


FIGURE 2. Network architecture of LCF-ATEPC model [34] for ABSA.

SemEval-2014 Task4, and it will be employed to perform ABSA later in this paper.

### E. REPUTATION GENERATION

Since we are dealing with opinions shared on the Internet, the reputation of an entity is formed and influenced by various factors that shift public opinion over time. These factors

include the popularity and reliability of the user sharing the online review, the time the review was posted online, and finally, the sentiment orientation of aspects of an entity within the online review. This section aims to introduce those three factors, and to calculate for each factor a numerical score that will be exploited to generate the final reputation value for a specific entity.

1) REVIEW POPULARITY SCORE

It is well known that social media and e-commerce platforms contain a huge amount of user reviews toward various online products. However, those reviews have different weights in affecting the opinion of other users as well as the reputation of the entity. A statistical study<sup>6</sup> was done in 2017 shows that 93% of consumers admit that online reviews influenced their purchase decisions, and 92% of consumers are more likely to purchase after reading a trusted review. In fact, what makes a review more influential than others is the amount of engagement it received from other users. Since we are dealing with reviews collected from multi-platforms on the Internet, we have chosen two types of engagement features: Likes and Shares. Likes represent a direct reflection of user preferences. If a user likes a review, he is more likely to be interested and approved of that review. However, by sharing a review, users essentially endorse the content to all of their followers.

The goal of this sub-section is to calculate a popularity score for each user’s review in order to differentiate between those reviews based on the amount of engagement received from likes and shares. As mentioned previously in the paper, our system collects data from two different types of platforms, the first one provides just the text and like features, the second type provides the text, like, and share features. For example, ‘TripAdvisor’ is considered as a platform from the first type, that doesn’t provide a share feature to the users, which means that the review can not be shared within the network, and it will have an assigned value of zero in our collected dataset for the attribute “review shares”. Therefore, we designed Equation (5) to calculate the popularity score for each review in our dataset. We multiplied the values of likes and shares in the equation with 0.5 in order to fit the calculated popularity score into an interval between 0 and 1. Notations used in this sub-section are listed in TABLE 3.

$$psr_{ij} = \frac{l_{ij} \times 0.5}{\max(L_j)} + \frac{s_{ij} \times 0.5}{\max(S_j)} \quad (5)$$

The result is a numerical value between 0 and 1 calculated for each review in order to indicate its popularity. A higher popularity score is an indicator that the review is influential. Those popularity scores are going to be used to compute the reputation for the target entity.

TABLE 3. Notions used in sub-section: Review popularity score.

Variable	Description
$psr_{ij}$	Popularity score of review $i$ expressed for an entity $j$
$l_{ij}$	Number of likes received for a review $i$ expressed for an entity $j$
$s_{ij}$	Number of shares received for a review $i$ expressed for an entity $j$
$\max(L_j)$	Max number of likes received by a review toward entity $j$
$\max(S_j)$	Max number of shares received by a review toward entity $j$

2) REVIEW TIME SCORE

One of the first things people do when reading a product review is to check the posting time of that review, and frequently focus on the most recent reviews. The date of the

reviews plays a major role in instilling confidence in potential customers since the ownership of a business can change, branding can change, and products and services are constantly evolving. Older reviews carry much less weight with consumers as shown in a recent study done by Brightlocal,<sup>7</sup> where about 85% of consumers consider any old review to be irrelevant, this can impact the computation of reputation for a product/service. However, time does not affect every product in every domain. An old review on the Internet about certain products like cheese or some type of old classic movies can still be relevant today. This means the date of the review doesn’t matter in this situation. In this sub-section, we proposed Equation (6) to calculate a review time score. The result of the proposed equation is a numerical score between 0 and 1 where most recent reviews will receive a score close to 1 and vice versa. Notations used in this sub-section are listed in TABLE 4. We mentioned that this feature will be optional in the proposed reputation system. The user can choose to discard the aspect of time, which means that the review time score will not be employed when generating the reputation for certain products or services.

$$tsr_{ij} = 1 - (y - rt_{ij}) \times 0.03 \quad (6)$$

TABLE 4. Notions used in sub-section: Review time score.

Variable	Description
$tsr_{ij}$	Time score of a review $i$ expressed for an entity $j$
$y$	Current year
$rt_{ij}$	Posting year of a review $r_{ijk}$

3) REVIEW SENTIMENT ANALYSIS

We employ the LCF-ATEPC model to the spam-free review dataset for the purpose of extracting the aspects from each review with their associated sentiment polarity. Then, we group the same extracted aspects for all the reviews with their sentiment polarity. The popularity score calculated previously for each review will also be assigned to its extracted aspects. TABLE 6 shows the procedure for extraction and sentiment prediction of each aspect within the review, while TABLE 7 shows the grouping of similar extracted aspects with their popularity scores.

TABLE 5. Notions used in sub-section: Review sentiment analysis.

Variable	Description
$asp_{ij}$	Aspect $i$ for entity $j$
$p_{ij}$	Total number of positive reviews toward $asp_{ij}$
$n_{ij}$	Total number of negative reviews toward $asp_{ij}$
$ssasp_{ij}$	Sentiment score toward aspect $asp_{ij}$

Our system compute a sentiment score  $ssasp_{ij}$  for each aspect  $asp_{ij}$  based on their positive and negative sentiment orientation using Equation 7. Notations used in this sub-section are listed in TABLE 5. Aspects with neutral sentiment

<sup>6</sup><https://www.podium.com/resources/podium-state-of-online-reviews/>

<sup>7</sup><https://www.brightlocal.com/research/local-consumer-review-survey/>



TABLE 6. Examples of the results obtained from the employment of the LCF-ATEPC on a sample of reviews.

Review	Aspect	Sentiment polarity	popularity score
The camera is good but the design is bad.	camera	Positive	0.4
	design	Negative	0.4
I love the camera and the design is so cool, but i have no opinion on the screen quality.	camera	Positive	0.2
	design	Positive	0.2
	screen	Neutral	0.2

TABLE 7. Example of similar aspects grouping step, post LCF-ATEPC employment.

Grouped Aspects	Sentiment polarities	popularity score
camera camera	Positive Positive	0.4 0.2
design design	Negative Positive	0.4 0.2
screen	Neutral	0.2

orientation will be dismissed and not going to be considered when calculating the  $ssasp_{ij}$ , since the user was sentimentally unbiased towards those aspects.

$$ssasp_{ij} = \frac{p_{ij}}{p_{ij} + n_{ij}} \tag{7}$$

F. REPUTATION COMPUTING

Based on the previously calculated features, the proposed system computes a reputation value for each aspect using Equation (8). The sentiment score  $ssasp_{ij}$  and the average time scores  $\frac{sum(T_{ij})}{m_{ij}}$  are multiplied by 9 in order to obtain a number bounded between 0 and 9, then we aggregate the result with a customized average of negative and positive popularity scores  $\frac{\sum PS_{pos_{ij}} - \sum PS_{neg_{ij}}}{L_{ij}}$ , which is bounded between 0 and 1. The final result is a numerical value between 0 and 10 that represents the reputation of an aspect  $asp_{ij}$ . Notations used in this sub-section are listed in TABLE 8.

$$Rep(asp_{ij}) = \max \left[ 1, \left( ssasp_{ij} \times \frac{sum(T_{ij})}{m_{ij}} \times 9 \right) + \frac{\sum PS_{pos_{ij}} - \sum PS_{neg_{ij}}}{L_{ij}} \right] \tag{8}$$

where

$$L_{ij} = \begin{cases} p_{ij}, & \text{if } \sum PS_{pos_{ij}} - \sum PS_{neg_{ij}} \geq 0 \\ n_{ij}, & \text{if } \sum PS_{pos_{ij}} - \sum PS_{neg_{ij}} \leq 0 \end{cases}$$

In order to generate the overall reputation for an entity, the system calculates the average of all aspects' reputation values using Equation (9).

$$FinalRep(E_j) = \frac{\sum_{i=0}^n Rep(asp_{ij})}{n} \tag{9}$$

TABLE 8. Notions used in sub-section: Reputation computing.

Variable	Description
$ssasp_{ij}$	Sentiment score toward aspect $asp_{ij}$
$T_{ij}$	Set of time scores of reviews that describe aspect $asp_{ij}$ .
$m_{ij}$	Number of reviews that describe aspect $asp_{ij}$ .
$PS_{pos_{ij}}$	Set of popularity scores of reviews that describe aspect $asp_{ij}$ positively.
$PS_{neg_{ij}}$	Set of popularity scores of reviews that describe aspect $asp_{ij}$ negatively.
$p_{ij}$	Total number of positive reviews toward $asp_{ij}$
$n_{ij}$	Total number of negative reviews toward $asp_{ij}$

G. REVIEW SCORE

The review score is a value calculated based on the popularity and time scores using Equation (10), and it is used to determine the most influential review. This score is not considered in the reputation value computation, and it is only employed during reputation visualisation in order to determine the most influential posting review.

$$rs_{ij} = \frac{psr_{ij} + tsr_{ij}}{2} \tag{10}$$

We denote:

$rs_{ij}$ : score of review  $i$  toward entity  $j$

H. REPUTATION VISUALIZATION

Our system provides an advanced efficiently-designed user-friendly interface compared with the previous reputation systems. The interface displays all details regarding the reputation of a specific product such as the overall reputation value of an entity, aspects reputation values, most reviewed aspects, and finally, the most influential reviews. As we can see in FIGURE 3, the proposed visualization tool is an interactive user interface where the user can obtain more detailed information about a specific feature, by placing a cursor over its position on the display using a pointing device, and it is then initiated by clicking. The proposed reputation visualization tool will help the users to have better insights toward the targeted product/service, and therefore supporting them during their decision-making process.

V. EXPERIMENT RESULTS

A. EXPERIMENTAL DATA COLLECTION AND PREPROCESSING

Four experimental review datasets were collected where each dataset belongs to a different domain (product, movie, hotel, restaurant). Every dataset contains a combination of reviews from various social media and e-commerce platforms, and each review includes a textual opinion expressed by the user, user name, review posting year, number of likes, number of shares, and the platform host. We hired four human



FIGURE 3. Reputation visualisation dashboard.

TABLE 9. Statistical information about the evaluation datasets.

Domain	Number of reviews	Host platforms
Movie	713	IMDb, Twitter, Facebook
Product	1109	Twitter, Amazon, Ebay
Hotel	686	TripAdvisor, Twitter, Facebook
Restaurant	725	Yelp, TripAdvisor, Twitter

annotators to manually label the four datasets by extracting and identifying the polarity of each aspect in the reviews. TABLE 9 shows the statistical information about the evaluation dataset while TABLE 10 shows review samples from one of those datasets. All the textual reviews of the dataset are cleaned and pre-processed by removing URLs, punctuations, special characters, and replacing slang words with formal ones. Finally, we prepare the cleaned textual reviews for the LCF-ATEPC model by performing tokenization and adding special tokens.

**B. OPINION SPAM DETECTION**

Due to a lack of availability of spam review datasets, an evaluation dataset with 1000 reviews was manually collected from various platforms on the internet. We hired annotators to manually annotate each user into two possible classes (Normal/Spammer) based on their review posting behavior. The outcome of this procedure results in identifying 682 genuine reviews and 318 spam reviews in our evaluation dataset. The spam review detection using behavioral features of the spammer has two phases: (1) calculating the spammer score  $Score(a)$  based on two spammer behavioral features  $CS$  and  $MNR$ . (2) Evaluating the performance of the

proposed spam review detection model by varying the value of the threshold from 0.50 to 0.68 with a step of 0.01, and using precision, recall, and accuracy as evaluation measures. TABLE 7 shows that the threshold value  $\tau = 0.57$  leads to the best performance in term of accuracy.

**C. ASPECT-BASED SENTIMENT ANALYSIS**

1) TRAINING DATASETS & HYPERPARAMETERS

LCF-ATEPC was trained on a merged dataset that we created from three most commonly used ABSA datasets, the Laptops and Restaurant datasets of SemEval-2014 Task4 and an ACL Twitter dataset [57]. The original three datasets were reformatted, and each sample was annotated with the Inside-outside-beginning (IOB) labels for ATE and polarity labels for APC tasks respectively. The polarity of each aspect may be positive, neutral, or negative. TABLE 12 depicts the details of these datasets. TABLE 13 shows the global hyperparameters settings for the model that leads to the best results according to the original authors of LCF-ATEPC model.

2) MODEL COMPARAISON

We have compared the LCF-ATEPC model with the following SOTA methods:

- AEN-BERT [58]: which is an attentional encoder network that employs the pre-trained BERT model to solve the APC.
- BERT-BASE [28]: which is the original pre-trained model. It was adapted to the aspect-based sentiment-analysis in order to automatically extract aspect terms and classify aspects' polarity.

TABLE 10. Extracted review samples from the “product” evaluation datasets.

Review text	User name	Number of likes	Number of shares	Review posting year	Platform
I am so excited to purchase this ...	@Marhiana33	23254	1509	2020	Twitter
Honestly, I would not waste a single penny on ...	Josh lu	3913	0	2021	Facebook

TABLE 11. The performances of the threshold-based detection method.

Threshold	Precision	Recall	Accuracy
0.50	0.689	0.681	0.686
0.51	0.683	0.675	0.673
0.52	0.701	0.697	0.70
0.53	0.712	0.702	0.709
0.54	0.728	0.719	0.723
0.55	0.743	0.741	0.739
0.56	0.827	0.820	0.822
0.57	0.898	0.883	0.891
0.58	0.735	0.726	0.733
0.59	0.737	0.731	0.728
0.60	0.714	0.712	0.711
0.61	0.714	0.712	0.714
0.62	0.714	0.706	0.712
0.63	0.702	0.696	0.699
0.64	0.708	0.701	0.706
0.65	0.675	0.669	0.661
0.66	0.668	0.666	0.659
0.67	0.581	0.581	0.572
0.68	0.585	0.584	0.579

TABLE 12. Description of the merged dataset including the SemEval-2014 Task & ACL 2014 Twitter.

Dataset	Positive		Negative		Neutral	
	Train	Test	Train	Test	Train	Test
Restaurant	2164	728	807	196	637	196
Laptop	994	341	870	128	464	169
Twitter	1561	173	1560	173	3127	346
Mixed	4719	1242	3237	497	4228	711

TABLE 13. Global hyperparameters settings exploited for the LCF-ATEPC, BERT-BASE, and BERT-SPC models in the experiments.

Hyperparameters	Settings
learning rate	$3 \times 10^{-8}$
batch size	16
training epochs	5
max sequence length	80
SRD	5
optimizer	AdamW

- BERT-SPC [58]: which is a fine-tuned BERT designed for text pair classification, and it is adapted to solve the aspect-based sentiment-analysis task.

TABLE 14 lists the main experimental results of LCF-ATEPC compared with the other ABSA-oriented models. Experimental results show that the LCF-ATEPC model achieves SOTA performance in both ATE and APC tasks.

### 3) LCF-ATEPC MODEL’S RESULTS ON THE EVALUATION DATASETS

The pre-trained LCF-ATEPC model was applied to the four evaluation datasets in order to assess the performance of ATE and APC. FIGURE 4 displays the F1-score obtained for

each dataset. We can observe that the model achieves good results in extracting the aspects and predicting their sentiment orientation. Interestingly enough, we can see that the model performed well on the movie and hotel datasets even though it was trained on the mixed dataset which doesn’t contain any reviews related to the previously mentioned domains.

### D. REPUTATION VISUALIZATION

Our proposed reputation generation system provides a detailed visualization concerning the reputation of a specific item. The output visual representation makes it easier to identify new insights about the target entity. The designed dashboard provides the reputation value of each aspect, the overall reputation of the entity, the most influential reviews as well as other statistical details. The Dashboards can be used by users in order to make data-driven business decisions. Compared with the previous reputation generation systems [3], [4], [6], our proposed system offers more advanced details about the reputation visualization of a specific entity. TABLE 15 displays the comparison results with the previous reputation generation systems concerning the visualization feature.

### E. SYSTEM EVALUATION

Previous reputation generation systems have focused on generating the reputation from either social media or e-commerce websites, relying on the overall sentiment of the collected reviews. In this paper, we proposed an advanced cross-platform system that:

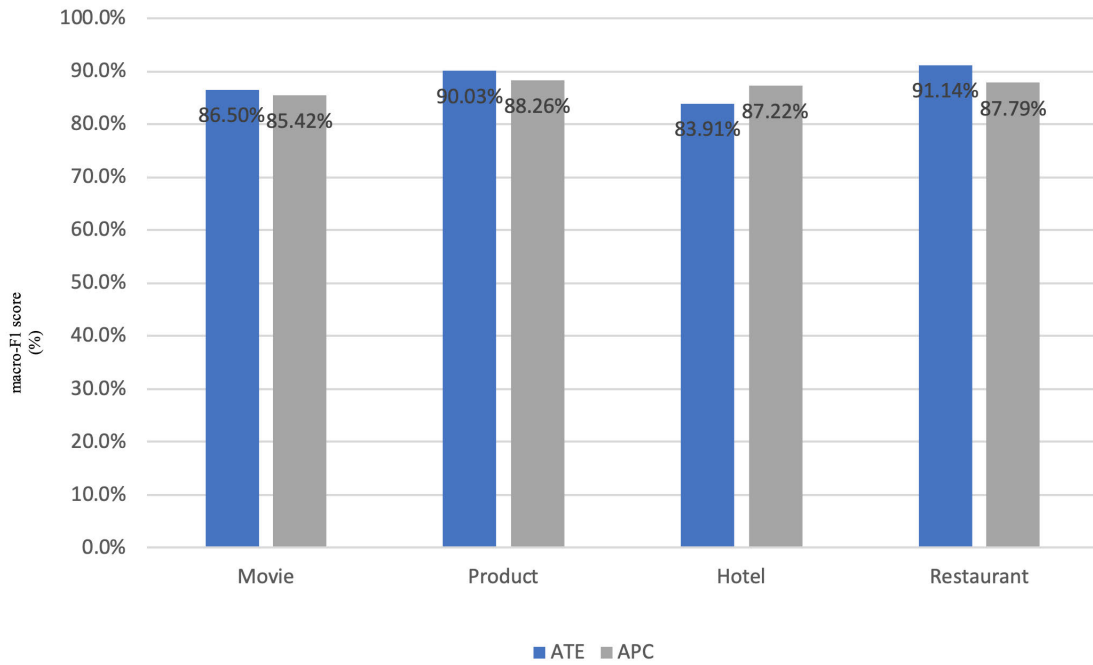
- 1) extracts and adjusts users’ data from various platforms at the same time, which allows it to generate a reliable reputation value.
- 2) incorporates a spam filtering mechanism to remove reviews written by potential spammers.
- 3) employs the aspect-based sentiment-analysis technique in order to extract aspects related to the target entity and predict their sentiment polarities, which allow us to generate the reputation of each aspect using mathematical formulas.
- 4) considers other important features such as the time feature and the popularity of the people sharing their opinions in order to increase the reliability of the generated reputation value.

TABLE 16 reveals the differences between previous reputation generation systems and our proposed system.

In order to evaluate the reliability of our system’s output and the effectiveness of its components, and due to the nonexistence of standard evaluation metrics for this kind of systems, we followed the same procedure used in

**TABLE 14.** Experimental results (%) of the LCF-ATEPC model.  $F1_{ate}$ ,  $Acc_{apc}$  and  $F1_{apc}$  are the macro-F1 score of aspect term extraction (ATE) subtask, accuracy and macro-F1 score of the aspect polarity subtask. The unreported experimental results are indicated by “-”. The “Δ” means the F1 score of the ATE task is not available for the BERT-SPC input format. The optimal performances are in Bold.

Models	Laptop			Restaurant			Twitter			Mixed		
	$F1_{ate}$	$Acc_{apc}$	$F1_{apc}$	$F1_{ate}$	$Acc_{apc}$	$F1_{apc}$	$F1_{ate}$	$Acc_{apc}$	$F1_{apc}$	$F1_{ate}$	$Acc_{apc}$	$F1_{apc}$
AEN-BERT	-	79.93	76.31	-	83.12	73.76	-	74.71	73.13	-	80.12	71.62
BERT-BASE	82.57	79.4	75.24	88.6	82.66	74.13	95.45	76.27	74.91	84.9	81.13	75.5
BERT-SPC	Δ	79.56	75.55	Δ	<b>86.77</b>	80.52	Δ	76.27	<b>75.16</b>	Δ	76.84	74.77
LCF-ATEPC	<b>83.82</b>	<b>80.97</b>	<b>77.86</b>	<b>89.02</b>	<b>86.77</b>	<b>80.45</b>	<b>96.4</b>	<b>76.7</b>	74.54	<b>85.44</b>	<b>83.02</b>	<b>76.18</b>



**FIGURE 4.** F1-score results for the ATE & APC tasks on the evaluation datasets.

**TABLE 15.** Comparison between the features displayed in the visualization tool among reputation generation systems.

System	Influential Reviews	Aspects reputation	Most reviewed Aspects
Yan et al.(2017) [3]	✗	✗	✗
Benlahbib and Nfaoui (2019) [4]	✗	✗	✗
Benlahbib and Nfaoui (2020) [6]	✓	✗	✗
This study	✓	✓	✓

**TABLE 16.** Comparison results: features exploited by reputation generation systems.

System	Review sentiment	Review popularity	Review posting year	Spam filtering	Aspect' reputation
Yan et al. (2017) [3]	✗	✗	✗	✓	✗
Benlahbib and Nfaoui (2019) [4]	✓	✗	✗	✗	✗
Benlahbib and Nfaoui (2020) [6]	✓	✓	✓	✗	✗
This study	✓	✓	✓	✓	✓

several works including [6], [59]. We have invited back the same 32 users from [6] that belong to different backgrounds (TABLE 17) for the reason of evaluating the effectiveness of 4 reputation generation systems: system 1 (our reputation system), system 2 [6], system 3 [4] and system 4 [3]. The volunteers were asked to assign a satisfaction score between 0 and 10 to each system based on its efficiency and helpfulness. To increase the validity of the experiments, we invited

three different experts to rate and judge each reputation system. TABLES 17 and 18 present information about the participants.

Each Volunteer rated the four systems based on their efficiency and helpfulness in supporting them during the decision making process while asking the question of “which system is more reliable and helpful?”. In TABLE 19, we calculated the average of all ratings provided by the users for each

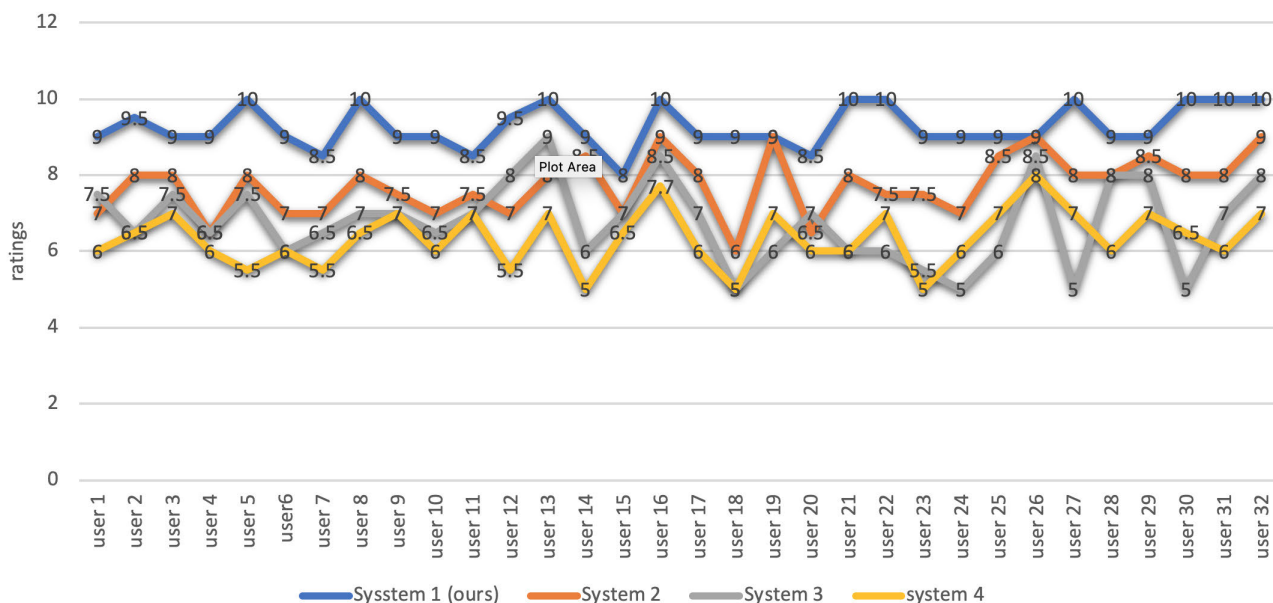


FIGURE 5. Comparison of users' ratings on SOTA reputation generation systems.

TABLE 17. Information about the 32 recruited participants for the assessment of the reputation generation systems.

Number of users	Background
6	Computer Science PhD students
2	Math PhD students
1	Electrical engineer
1	Undergraduate math student
2	computer science engineers
1	Physics teacher
4	Math teachers
1	Research engineer in computer science
1	Electronic engineer
1	Information system engineer
1	Student at the national school of commerce and management
1	Quality control technician
7	Medical student
1	Housewife

system  $\mu = \frac{1}{n} \sum_{i=0}^n x$ , where  $x_1, x_2, \dots, x_n$  are the observed ratings and  $n$  is the total number of ratings. We also measured the coefficient of variation (CV), which is the standard deviation divided by the mean times 100% as shown in Equation 11.

$$CV = \frac{\sigma}{\mu} 100\% \tag{11}$$

where

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$$

We denote:

$\sigma$ : Standard deviation

As we can see, our system was higher-rated among the other systems based on the average rating. Moreover, our proposed system is the only one to get the perfect rating (10 out of 10) from 10 different users. With an average rating of 9.33, our proposed system is ranked first in comparison with the others. System 2 got the second higher rating

TABLE 18. Information about the 3 recruited experts for the assessment of the reputation generation systems.

Expert	Background
Expert 1	Owner and reputation manager of the streaming website of 'Fushar.tv'
Expert 2	Social media manager of an online store
Expert 3	Computer science professor specialized in business intelligence

with 7.78. In third place, we have system 3 with an average rating of 6.63. Finally, system 4 came in last with an average rating of 6.37, which is close to system 3 since both systems only use the sentiment or semantic features for generating the reputation on an entity. FIGURE 5 displays the ratings given by the users.

We also calculated and compared the coefficient of variation of the users' ratings for each system in order to measure the spread of the ratings. If the ratings all lie close to the mean/average, then the percentage of the coefficient of variation will be small, while if the ratings are spread out over a large range of values, then the percentage coefficient of variation will be large. This will help us determine if the ratings given for each system are balanced. As we can see in TABLE 19, the coefficient of variation of the group of ratings for our proposed system is the lowest compared to the other systems with a value of 6.12%.

In addition to the 32 voluntary users, we also asked three experts (TABLE 18) to rate the four systems based on their helpfulness and functionality. The results are shown in TABLE 20. As we can see, all the experts favor our proposed reputation generation system by giving it a higher rating score compared with the other systems. With an average rating of '8.83', our proposed system takes first place, then system 2 in second place with an average rating of '7.5', next is system 3 in third place with an average score of '6.83', and finally system 4 in the last place with an average rating

TABLE 19. User satisfaction comparison.

	System 1 (ours)	System 2 [6]	System 3 [4]	System 4 [3]
User 1	<b>9</b>	7	7.5	6
User 2	<b>9.5</b>	8	6.5	6.5
User 3	<b>9</b>	8	7.5	7
User 4	<b>9</b>	6.5	6.5	6
User 5	<b>10</b>	8	7.5	5.5
User 6	<b>9</b>	7	6	6
User 7	<b>8.5</b>	7	6.5	5.5
User 8	<b>10</b>	8	7	6.5
User 9	<b>9</b>	7.5	7	7
User 10	<b>9</b>	7	6.5	6
User 11	<b>8.5</b>	7.5	7	7
User 12	<b>9.5</b>	7	8	5.5
User 13	<b>10</b>	8	9	7
User 14	<b>9</b>	8.5	6	5
User 15	<b>8</b>	7	7	6.5
User 16	<b>10</b>	9	8.5	7.7
User 17	<b>9</b>	8	7	6
User 18	<b>9</b>	6	5	5
User 19	<b>9</b>	9	6	7
User 20	<b>8.5</b>	6.5	7	6
User 21	<b>10</b>	8	6	6
User 22	<b>10</b>	7.5	6	7
User 23	<b>9</b>	7.5	5.5	5
User 24	<b>9</b>	7	5	6
User 25	<b>9</b>	8.5	6	7
User 26	<b>9</b>	<b>9</b>	8.5	8
User 27	<b>10</b>	8	5	7
User 28	<b>9</b>	8	8	6
User 29	<b>9</b>	8.5	8	7
User 30	<b>10</b>	8	5	6.5
User 31	<b>10</b>	8	7	6
User 32	<b>10</b>	9	8	7
<b>Average</b>	<b>9,33</b>	7,78	6,63	6,37
<b>Coefficient of variation</b>	<b>6.12%</b>	10.12%	16.10%	11.90%

TABLE 20. Experts satisfaction comparison.

	System 1 (ours)	System 2	System 3	System 4
Expert 1	<b>9</b>	8	7	6.5
Expert 2	<b>8.5</b>	7	6.5	5.5
Expert 3	<b>9</b>	7.5	7	6
<b>Average</b>	<b>8.83</b>	7.5	6.83	6.0

of '6.0'. FIGURE 6 shows a comparison between users' and experts' average ratings, which indicates that our system is more reliable to generate and visualize reputation compared

with the previous systems. We also asked the experts to share a review where they expressed their opinions about the proposed system, which is presented in TABLE 21.

## VI. DISCUSSION

The system proposed in this paper can be defined as an advanced decision-making tool, capable of producing numerical values that reflect the reputation of an entity (products, services, movies, hotels, etc.) from opinions and reviews shared on the internet. The proposed system is the first to deal

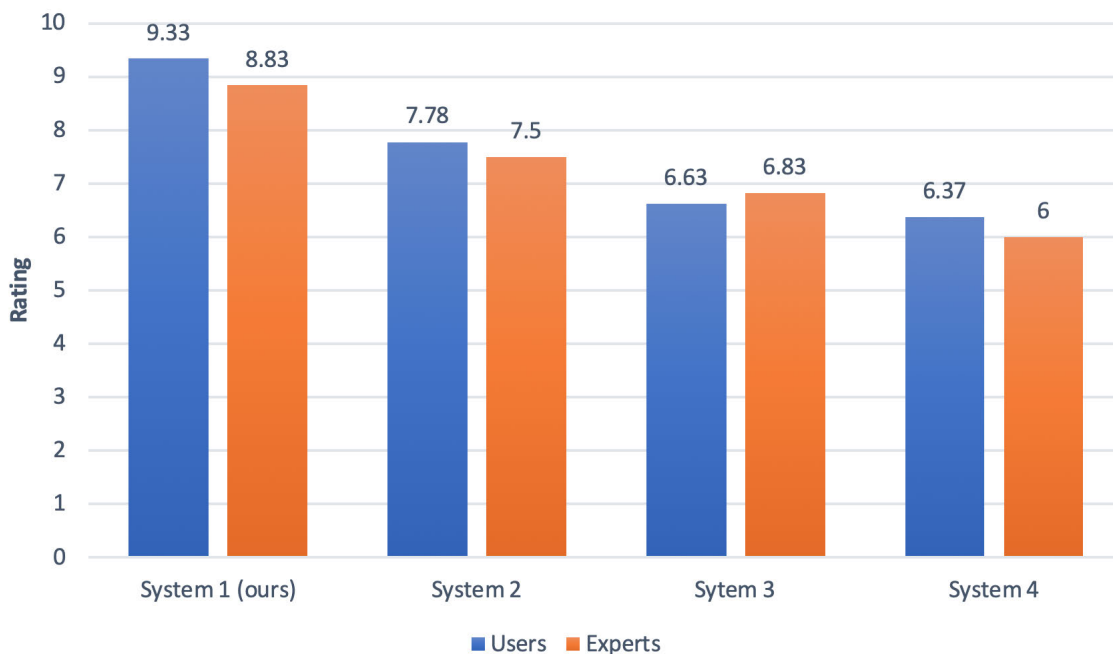


FIGURE 6. Comparison between users' and experts' average ratings on SOTA reputation generation systems.

TABLE 21. Experts reviews.

Expert	Review
Expert 1	As an online service provider, spam reviews are a major obstacle to a successful understanding of the viewpoint of users toward our service. System 1 provides the solution for this problem as well as employing other extremely important features such as processing reviews from different platforms and more. For those previous reasons, we believe that System 1 is an improved version compared with the other systems in terms of generating a trusted reputation report.
Expert 2	System 1 is determined as the perfect tool for small/big e-commerces businesses to understand the quality and characteristics of their output. Providing a dashboard that displays the aspects' reputation of a product will definitely improve its quality. Also, taking into consideration the aspect of time while generating the reputation is an important feature to have in a reputation generation system.
Expert 3	System 1 does an excellent job of incorporating various technologies for the purpose of generating the reputation of an item. I offered system 1 the higher rating '9/10' since it uses the ABSA to generate the reputation score for each aspect which is unfortunately not present in other systems. By exploiting other features such as popularity score, time score, and the spam filter, I believe System 1 is able to generate a more reliable reputation value compared with the other systems.

with opinions from various platforms by having the ability to process features of different platforms with high flexibility. The proposed reputation system is also the first to integrate an opinion spam filter, that detects and eliminates spam opinions based on the characteristics of spammers' behaviors, making our system more secure from spammers attacks, which leads to the generation of reliable reputation values. Furthermore, one of its main components is the ability to extract and

analyze aspects of the target entity using SOTA aspect-based sentiment-analysis tools. The system additionally incorporates the opinions posting time and popularity features for the purpose of generating reputation, which makes it more reliable and trustworthy. A visualization tool is proposed, where the detailed output results of the whole reputation generation procedure are displayed in an interactive user-friendly interface, which will facilitate the online decision-making process for both regular users and business owners.

### VII. CONCLUSION

In this paper, we proposed a reputation system capable of generating numerical reputation values for a specific item (product, movie, service, hotel, etc.) and its aspects based on opinions and reviews expressed online. The contribution of this work revolves around four components that were not exploited in previous systems. The first one is cross-platform compatibility, where the proposed system can collect and process opinions from different platforms (Facebook, Amazon, Twitter, TripAdvisor, etc.) as well as managing and standardizing those platforms' features. The second one is opinion spam filtering, where the spam opinions are detected and eliminated based on spammers' behavior features, keeping only authentic opinions. The third one is employing a SOTA aspect-based sentiment-analysis model named LCF-ATEPC in order to extract and analyze the aspects within the textual opinions. Finally, we incorporated the previous results with a calculated review time score and review popularity score using mathematical formulas to obtain a reputation value for the targeted entity as well as the reputation values of the entities' aspects. In addition,

a holistic reputation visualization is provided within the system that displays the detailed output results of the reputation generation process. To assess the effectiveness of our reputation system, we invited 32 participants and 3 experts to choose the best performing system out of four SOTA reputation systems by giving numerical satisfaction scores to each system. Our reputation system achieved the highest average satisfaction scores from both users and experts. In the future, we propose to investigate the effectiveness of our proposed system by attempting to generate more than the numerical reputation values, such as extending the system to automatically generate a textual summary of the benefits and drawbacks of the targeted entity. Also, we intend to extend this system to be used in multilingual content.

## REFERENCES

- [1] Abdel-Hafez, Y. Xu, and D. Tjondronegoro, "Product reputation model: An opinion mining based approach," in *Proc. 1st Int. Workshop Sentiment Discovery Affect. Data*, vol. 917, London, U.K., Jun. 2013, pp. 16–27. [Online]. Available: <https://eprints.qut.edu.au/58118/>
- [2] U. Farooq, A. Nongailard, Y. Ouzrout, and M. A. Qadir, "A feature-based reputation model for product evaluation," *Int. J. Inf. Technol. Decis. Making*, vol. 15, no. 6, pp. 1521–1553, Nov. 2016, doi: [10.1142/S0219622016500358](https://doi.org/10.1142/S0219622016500358).
- [3] Z. Yan, X. Jing, and W. Pedrycz, "Fusing and mining opinions for reputation generation," *Inf. Fusion*, vol. 36, pp. 172–184, Jul. 2017, doi: [10.1016/j.inffus.2016.11.011](https://doi.org/10.1016/j.inffus.2016.11.011).
- [4] A. Benlabib and E. H. Nfaoui, "A hybrid approach for generating reputation based on opinions fusion and sentiment analysis," *J. Organizational Comput. Electron. Commerce*, vol. 30, no. 1, pp. 9–27, 2020, doi: [10.1080/10919392.2019.1654350](https://doi.org/10.1080/10919392.2019.1654350).
- [5] E. I. Elmurngi and A. Gherbi, "Building sentiment analysis model and compute reputation scores in E-commerce environment using machine learning techniques," *Int. J. Organizational Collective Intell.*, vol. 10, no. 1, pp. 32–62, Jan. 2020.
- [6] A. Benlabib and E. H. Nfaoui, "Aggregating customer review attributes for online reputation generation," *IEEE Access*, vol. 8, pp. 96550–96564, 2020, doi: [10.1109/ACCESS.2020.2996805](https://doi.org/10.1109/ACCESS.2020.2996805).
- [7] A. Gupta, S. Priyani, and R. Balakrishnan, "Customized reputation generation of entities using sentiment analysis," *World J. Eng.*, vol. 18, no. 4, pp. 596–605, Jul. 2021, doi: [10.1108/WJE-09-2020-0470](https://doi.org/10.1108/WJE-09-2020-0470).
- [8] A. Boumhidi and E. Nfaoui, "Leveraging lexicon-based and sentiment analysis techniques for online reputation generation," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 6, pp. 274–289, 2021, doi: [10.22266/ijies2021.1231.25](https://doi.org/10.22266/ijies2021.1231.25).
- [9] V. M. Pradhan, J. Vala, and P. Balani, "A survey on sentiment analysis algorithms for opinion mining," *Int. J. Comput. Appl.*, vol. 133, no. 9, pp. 7–11, Jan. 2016.
- [10] A. Tripathy, A. Anand, and S. K. Rath, "Document-level sentiment classification using hybrid machine learning approach," *Knowl. Inf. Syst.*, vol. 53, no. 3, pp. 805–831, 2017.
- [11] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing*, 2nd ed. Boca Raton, FL, USA: Taylor & Francis, 2010.
- [12] K. Schouten and F. Frasinca, "Survey on aspect-level sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, Mar. 2016.
- [13] A. Da'u and N. Salim, "Aspect extraction on user textual reviews using multi-channel convolutional neural network," *PeerJ Comput. Sci.*, vol. 5, p. e191, May 2019.
- [14] C. Feng, Y. Rao, A. Nazir, L. Wu, and L. He, "Pre-trained language embedding-based contextual summary and multi-scale transmission network for aspect extraction," *Proc. Comput. Sci.*, vol. 174, pp. 40–49, Jan. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050920315672>
- [15] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: A hybrid network for targeted aspect-based sentiment analysis," *Cogn. Comput.*, vol. 10, pp. 639–650, Mar. 2018.
- [16] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2004, pp. 168–177, doi: [10.1145/1014052.1014073](https://doi.org/10.1145/1014052.1014073).
- [17] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann, 1994, pp. 487–499. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645920.672836>
- [18] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowl.-Based Syst.*, vol. 108, pp. 42–49, Sep. 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950705116301721>
- [19] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Austin, TX, USA, Nov. 2016, pp. 606–615. [Online]. Available: <https://aclanthology.org/D16-1058>
- [20] S. Hochreiter and J. J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [21] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, Dublin, Ireland, 2014, pp. 27–35. [Online]. Available: <https://aclanthology.org/S14-2004>
- [22] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, Melbourne, VIC, Australia, Vol. 1, Jul. 2018, pp. 2514–2523. [Online]. Available: <https://aclanthology.org/P18-1234>
- [23] Y. LeCun and Y. Bengio, *Convolutional Networks for Images, Speech, and Time Series*. Cambridge, MA, USA: MIT Press, 1998, pp. 255–258.
- [24] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An interactive multi-task learning network for end-to-end aspect-based sentiment analysis," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, Florence, Italy, 2019, pp. 504–515. [Online]. Available: <https://aclanthology.org/P19-1048>
- [25] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 task 12: Aspect based sentiment analysis," in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, Denver, CO, USA, 2015, pp. 486–495. [Online]. Available: <https://aclanthology.org/S15-2082>
- [26] L. Li, Y. Liu, and A. Zhou, "Hierarchical attention based position-aware network for aspect-level sentiment analysis," in *Proc. 22nd Conf. Comput. Natural Lang. Learn.*, Brussels, Belgium, 2018, pp. 181–189. [Online]. Available: <https://aclanthology.org/K18-1018>
- [27] H. Xu, B. Liu, L. Shu, and P. Yu, "BERT post-training for review reading comprehension and aspect-based sentiment analysis," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, Minneapolis, MN, USA, vol. 1, Jun. 2019, pp. 2324–2335. [Online]. Available: <https://aclanthology.org/N19-1242>
- [28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, Minneapolis, MN, USA, vol. 1, Jun. 2019, pp. 4171–4186. [Online]. Available: <https://aclanthology.org/N19-1423>
- [29] A. Karimi, L. Rossi, and A. Prati, "Adversarial training for aspect-based sentiment analysis with BERT," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 8797–8803.
- [30] A. Rietzler, S. Stabinger, P. Opitz, and S. Engl, "Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification," in *Proc. 12th Lang. Resour. Eval. Conf.*, Marseille, France, May 2020, pp. 4933–4941. [Online]. Available: <https://aclanthology.org/2020.lrec-1.607>
- [31] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2019. [Online]. Available: <https://proceedings.neurips.cc/paper/2019/file/dc6a7e655d7e5840e66733e9ee67cc69-Paper.pdf>
- [32] J. Dai, H. Yan, T. Sun, P. Liu, and X. Qiu, "Does syntax matter? A strong baseline for aspect-based sentiment analysis with RoBERTa," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, Jun. 2021, pp. 1816–1829. [Online]. Available: <https://aclanthology.org/2021.naacl-main.146>
- [33] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.
- [34] H. Yang, B. Zeng, J. Yang, Y. Song, and R. Xu, "A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction," *Neurocomputing*, vol. 419, pp. 344–356, Jan. 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231220312534>



- [35] B. Zeng, H. Yang, R. Xu, W. Zhou, and X. Han, "LCF: A local context focus mechanism for aspect-based sentiment classification," *Appl. Sci.*, vol. 9, no. 16, p. 3389, Aug. 2019. [Online]. Available: <https://www.mdpi.com/2076-3417/9/16/3389>
- [36] R. Paul and Z. Richard, *Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System*. Bingley, U.K.: Emerald Group, 2002.
- [37] J. Schneider, G. Kortuem, J. Jager, S. Fickas, and Z. Segall, "Disseminating trust information in wearable communities," *Pers. Technol.*, vol. 4, no. 4, pp. 245–248, Jan. 2000. [Online]. Available: <http://dx.doi.org/10.1007/PL00000012>
- [38] F. Garcin, B. Faltings, and R. Jurca, "Aggregating reputation feedback," in *Proc. 1st Int. Conf. Reputation: Theory Technol.*, vol. 1, 2009, pp. 62–74.
- [39] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara, "Reputation systems," *Commun. ACM*, vol. 43, no. 12, pp. 45–48, 2000.
- [40] C. S. Leberknight, S. Sen, and M. Chiang, "On the volatility of online ratings: An empirical study," in *E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life*, M. J. Shaw, D. Zhang, and W. T. Yue, Eds. Berlin, Germany: Springer, 2012, pp. 77–86.
- [41] J. Cho, K. Kwon, and Y. Park, "Q-rater: A collaborative reputation system based on source credibility theory," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3751–3760, Mar. 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417408001486>
- [42] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decis. Support Syst.*, vol. 43, no. 2, pp. 618–644, Mar. 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167923605000849>
- [43] A. Boumhidi and E. H. Nfaoui, "Mining user's opinions and emojis for reputation generation using deep learning," in *Proc. 4th Int. Conf. Intell. Comput. Data Sci. (ICDS)*, 2020, pp. 1–5.
- [44] A. Boumhidi, A. Benlahbib, and E. H. Nfaoui, "Mining online opinions and reviews using bi- lstm for reputation generation," in *Artificial Intelligence and Industrial Applications*. Cham, Switzerland: Springer, 2021, pp. 135–144.
- [45] A. Benlahbib and E. Nfaoui, "MTVRep: A movie and tv show reputation system based on fine-grained sentiment and semantic analysis," *Int. J. Elect. Comput. Eng.*, vol. 11, no. 2, pp. 1613–1626, 2021, doi: [10.11591/ijece.v11i2.pp1613-1626](https://doi.org/10.11591/ijece.v11i2.pp1613-1626).
- [46] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNLP natural language processing toolkit," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics: Syst. Demonstrations*. Baltimore, MD, USA, Jun. 2014, pp. 55–60. [Online]. Available: <https://www.aclweb.org/anthology/P14-5010>
- [47] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, New Orleans, LA, USA, vol. 1, Jun. 2018, pp. 2227–2237. [Online]. Available: <https://www.aclweb.org/anthology/N18-1202>
- [48] A. Esuli and F. Sebastiani, "SENTIWORDNET: A publicly available lexical resource for opinion mining," in *Proc. 5th Int. Conf. Lang. Resour. Eval. (LREC)*. Genoa, Italy, May 2006, pp. 1–6. [Online]. Available: <http://www.lrec-conf.org/proceedings/lrec2006/pdf/384.pdf>
- [49] S. Baccianella, A. Esuli, and F. Sebastiani, "SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proc. 7th Conf. Int. Lang. Resour. Eval. (LREC)*, Valletta, Malta, May 2010, pp. 1–5. [Online]. Available: [http://www.lrec-conf.org/proceedings/lrec2010/pdf/769\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf)
- [50] A. Benlahbib and E. H. Nfaoui, "An unsupervised approach for reputation generation," *Proc. Comput. Sci.*, vol. 148, pp. 80–86, Jan. 2019, doi: [10.1016/j.procs.2019.01.011](https://doi.org/10.1016/j.procs.2019.01.011).
- [51] A. Benlahbib, A. Boumhidi, and E. H. Nfaoui, "A logistic regression approach for generating movies reputation based on mining user reviews," in *Proc. Int. Conf. Intell. Syst. Adv. Comput. Sci. (ISACS)*, Dec. 2019, pp. 1–7, doi: [10.1109/ISACS48493.2019.9068916](https://doi.org/10.1109/ISACS48493.2019.9068916).
- [52] A. Benlahbib and E. H. Nfaoui, "AmazonRep: A reputation system to support Amazon's customers purchase decision making process based on mining product reviews," in *Proc. 4th Int. Conf. Intell. Comput. Data Sci. (ICDS)*, Oct. 2020, pp. 1–8.
- [53] J. G. Biradar, "Exponential distribution model for review spam detection," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 3, pp. 938–947, Apr. 2017. Accessed: Dec. 31, 2021. [Online]. Available: <http://www.ijarcs.info/index.php/Ijarcs/article/view/3129>, doi: [10.26483/ijarcs.v8i3.3129](https://doi.org/10.26483/ijarcs.v8i3.3129).
- [54] I. Dematis, E. Karapistoli, and A. Vakali, "Fake review detection via exploitation of spam indicators and reviewer behavior characteristics," in *Proc. Int. Conf. Current Trends Theory Pract. Inform.* Springer, 2018, pp. 581–595.
- [55] N. Hussain, H. Mirza, and I. Hussain, "Detecting spam review through spammer's behavior analysis," *Adv. Distrib. Comput. Artif. Intell. J.*, vol. 8, p. 61, Mar. 2019.
- [56] N. Hussain, H. Mirza, and I. Hussain, "Detecting spam review through spammer's behavior analysis," *Adv. Distrib. Comput. Artif. Intell. J.*, vol. 8, p. 61, Mar. 2019.
- [57] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of Twitter data," in *Proc. Workshop Lang. Social Media (LSM)*. Portland, OR, USA, Jun. 2011, pp. 30–38. [Online]. Available: <https://aclanthology.org/W11-0705>
- [58] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Attentional encoder network for targeted sentiment classification," 2019, *arXiv:1902.09314*.
- [59] D. Wang, S. Zhu, and T. Li, "SumView: A web-based engine for summarizing product reviews and customer opinions," *Expert Syst. Appl.*, vol. 40, no. 1, pp. 27–33, Jan. 2013, doi: [10.1016/j.eswa.2012.05.070](https://doi.org/10.1016/j.eswa.2012.05.070).



**ACHRAF BOUMHIDI** is currently pursuing the Ph.D. degree with the Faculty of Science Dhar El Mahraz, Fez, Morocco. His research interests include surrounding natural language processing (NLP) and social network analysis (SNA) for decision making in social media platforms.



**ABDESSAMAD BENLAHBIB** received the Ph.D. degree in computer science. He has published several papers in journals and conferences in the area of computer science, such as IEEE Access, *Journal of Organizational Computing and Electronic Commerce*, *International Journal of Electrical and Computer Engineering*, and *SemEval*. His research interest includes the application of natural language processing techniques to support customers during their decision-making process in e-commerce platforms.



**EL HABIB NFAOUI** (Member, IEEE) received the joint Ph.D. degree in computer science from the University of Sidi Mohamed Ben Abdellah, Morocco, and the University of Lyon, France, in 2008, under a Cotutelle Agreement (doctorate in joint-supervision), and the HU Diploma degree (accreditation to supervise research) in computer science from the University of Sidi Mohamed Ben Abdellah, in 2013. He is currently a Professor of computer science with the University of Sidi Mohamed Ben Abdellah, Fez, Morocco. He has published in international reputed journals, books, and conferences, and has edited seven conference proceedings and special issue books. His current research interests include information retrieval, language representation learning, machine learning and deep learning, web mining and text mining, semantic web, web services, social networks, and multi-agent systems. He is a Co-Founder and the Chair of the IEEE Morocco Section Computational Intelligence Society Chapter. He is a Co-Founder and an Executive Member of the International Neural Network Society Morocco Regional Chapter. He co-founded the International Conference on Intelligent Computing in Data Sciences (ICSD2017) and the International Conference on Intelligent Systems and Computer Vision (ISCV2015). He has served as a reviewer for scientific journals and as program committee of several conferences.