Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-16293-x

Aspect-based classification method for review spam detection

Mengsi Cai[1] · **Yonghao Du**[1] · **Yuejin Tan**[1] · **Xin Lu**[1]

Received: 28 September 2022 / Revised: 29 June 2023 / Accepted: 10 July 2023

Abstract
Online reviews have become available for consumers' reference to make purchase decisions, but a large number of spam reviews have damaged e-commerce reputations. Previous research has addressed review spam detection with classification models using textual features, behavior features, and relational features. However, the fine-grained aspect features related to the product attributes in online reviews have been overlooked and have not yet been thoroughly studied. Therefore, this study proposes a review spam detection model based on a list of novel aspect features. The basic idea is that since spam reviews are usually written by users without real experience, the product aspects depicted in spam reviews will be different from those in genuine reviews. First, we use the Bi-LSTM model to automatically extract massive aspect words, which are then clustered into different aspect categories by the K-means algorithm. Further, we propose nine novel aspect features to train a machine learning model for review spam detection. Experimental results on two labeled Yelp datasets show that the proposed aspect features can significantly improve the accuracy of review spam detection by about 16.11% to 38.86% compared with textual and behavior features.

Keywords Aspect features · Aspect extraction · Spam reviews · Review spam detection

1 Introduction

With the rapid growth and development of the e-commerce market, a vast amount of product reviews are generated online [5]. For example, millions of user-submitted reviews for various stores and products have been posted to online review sites such as Yelp.com and ResellerRatings.com. Since online reviews contain highly valuable information about the quality of products and services, people are increasingly depending on them to search for product information and make purchase decisions [7, 49]. Moreover, online reviews are particularly important for user requirement elicitation and new feature development from a manufacturer's perspective [6, 61].

✉ Xin Lu
xin.lu.lab@outlook.com

[1] College of Systems Engineering, National University of Defense Technology, Changsha 410073, China

However, unfortunately, a substantial amount of reviews are produced with deceptive opinions [17]. Such reviews for promoting or demoting certain products are called review spam or fake reviews and the persons publishing spam reviews are called spammers [33]. It is estimated that review spam accounts for 2–6% of websites like Orbitz, Priceline, Expedia, and TripAdvisor [39]. Also, it has been reported that more than 33% of reviews on the Internet are spam reviews and this proportion is increasing [56]. Therefore, detecting spam reviews has become a big concern in present times to authenticate online opinions and build trust with consumers [21].

Prior studies have been carried out to develop review spam classification models using textual features, behavior features, and relation-based features [37]. However, it is well recognized that there should be a latent but overlooked feature to capture the characteristics of spam reviews, i.e., aspect ratings which refer to users' attitudes to particular product attributes. For this purpose, Gao et al. [12] extracted the statistical features of movie reviews, which revealed that users express their sentiments on different aspects of movies in reviews. You et al. [58] proposed an aspect-rating local outlier factor model (AR-LOF) to identify spam reviews, in which an aspect segmentation algorithm proposed by Wang et al. [51] was adopted for extracting aspect units from the review texts. Xue et al. [54] developed a detection scheme based on the deviation of aspect-specific ratings between individual reviews and whole reviews. Nevertheless, previous studies mainly focused on the aspect ratings for identifying spam reviews, other aspect-related features have rarely been considered. Also, they defined aspect units in a manual or semi-automatic manner, resulting in a limited number of aspects for review spam detection.

To address the above problems, we propose an aspect-based classification method called ABCM for review spam detection. ABCM aims to identify spam reviews and non-spam reviews by automatically extracting a list of novel fine-grained aspect features. The basic idea is that since spammers may not have real experiences of purchasing or using the product, the aspects depicted in spam reviews will be different from those in genuine reviews [58]. If detailed information about product aspects is extracted from the review texts, then it is possible to differentiate spam reviews from non-spam reviews. The proposed approach is composed of three main steps: 1) aspect extraction, which automatically extracts a large number of aspect words and related user opinion words from the review texts through the Bi-LSTM (bidirectional long short-term memory) model, and then classifies the extracted aspect words into different aspect categories via the K-means clustering model, 2) feature extraction, which defines a list of novel aspect features that have never been used for review spam detection, and 3) classification, which identifies spam reviews based on an integrated supervised machine learning model XGBoost (eXtreme Gradient Boosting).

The main contributions of this paper are summarized as follows:

1 We extract a large number of product aspects with a suitable categorization from massive online reviews in an automatic manner, which helps understand user opinions towards the fine-grained product attributes;
2 We propose a list of novel aspect features that have never been used for review spam detection in previous studies, which are good at capturing potential spamming clues hidden in the review texts from a more micro perspective;
3 We develop an aspect-based classification method consisting of aspect extraction, feature extraction, and classification modules, which detects spam reviews with significantly improved performance than existing methods in previous studies on the same datasets;

4 We reveal the characteristics of spam reviews that are different from genuine reviews, which can be used by consumers to identify spam reviews in real life.

The rest of this paper is organized as follows. Section 2 reviews the literature relevant to this work. Section 3 introduce ABCM for review spam detection. In Sect. 4, we perform comparative studies between ABCM and other state-of-the-art approaches based on two Yelp datasets. Evaluation results are presented in Sect. 5, we also discuss the computational complexity of ABCM, the effect of aspect settings, the implication of aspect features, and evaluate the performance of ABCM on two Amazon datasets. Finally, Sect. 6 concludes the whole paper.

2 Related work

Since the seminal work by Jindal and Liu [20, 21], review spam detection has received great attention among both practitioners and researchers. A large number of textual (linguistic) features, behavior (structural) features, and relational (graph-based) features have been proposed for detecting spam reviews in previous studies [37].

Textual features imply valuable information contained in the text of reviews, including *n*-gram features [35, 39, 48, 54], Linguistic Inquiry and Word Count (LIWC) outputs [16], parts of speech (POS) tag frequencies [27], syntactic features [11, 45], and semantic features [22, 59]. Instead of manually designing discrete textual features, convolutional networks were used to learn the document-level review representation for review spam detection [4, 28]. Similarly, Yuan et al. [59] developed a hierarchical fusion attention network to automatically learn the semantics of reviews from the user and product level, which helps capture the complex semantics of reviews.

Behavior features refer to behavior information and metadata of review activities, including rating-based features (e.g., rating, extreme rating, rating deviation) [1, 29, 32] and time-related features (e.g., time of review, early time review, burst characteristics) [10, 23, 24, 34, 50]. Besides, the personal characteristics of users were also used, such as email ID, geographical location, and IP address [2]. Behavior features were widely used for identifying spam reviews and produced encouraging results. Many studies revealed that submitting duplicate or near-duplicate reviews on the same product is an abnormal behavior [18, 53]. In addition, the reviews intensely posted in a short time period with burstiness pattern have high probability to be spam reviews [24].

Considering that the textual and behavior features might be manipulated by spammers, relational features are extracted from the inter- and intra-relationships that exist among review entities like products, reviews, and reviewers [1, 30, 38]. For example, Li et al. [28] integrated a heterogeneous graph and a homogeneous graph to capture the local context and global context of a review. Furthermore, Shehnepoor et al. [44] utilized spam features as heterogeneous information networks to map the spam detection procedure into a classification problem in such networks. Based on the review graphs, a loopy belief propagation (LBP) algorithm was widely used to infer the final probabilities of different reviews being fake [41, 43].

In recent years, researchers have tried to investigate online reviews from the aspect perspective [31]. Noting that users usually comment on product attributes (i.e., aspects) in online reviews, their opinions on particular aspects were used to infer the plain-text

feedback of product ratings [55]. Compared with the numerical rating of overall performance (e.g. on a scale of 1–5), the detailed feedback on product aspects provides valuable experiences and opinions for others to make purchase decisions. Recent studies also revealed that the fine-grained aspects depicted in spam reviews are different from those in genuine reviews [12, 54, 58], which makes aspect ratings become a valuable feature for identifying spam reviews.

Based on the various features of review entities, machine learning techniques are the most promising methods for identifying spam reviews and non-spam reviews [40]. Popular supervised methods like support vector machine (SVM) [9, 22, 36], random forest [8], neural networks [28, 42, 43, 60], and logistic regression [18, 19] were employed to train classifiers on the labeled review datasets such as Yelp datasets [24, 43], Amazon Mechanical Turk (AMT) datasets [30, 39], and TripAdvisor datasets [27]. On the other hand, due to the lack of ground-truth datasets [52], many unsupervised or semi-supervised methods like co-training framework [25], cold start framework [50, 47], and PU (positive-unlabeled) learning [14, 15, 26] were conducted to overcome the data labeling challenges. However, existing studies have concentrated on single machine learning models and simple aspect-related features, in this paper, we aim to propose a list of novel aspect-based features for review spam detection and employ an ensemble learning model to improve classification performance.

3 Method

The framework of the proposed ABCM for review spam detection is illustrated in Fig. 1. Firstly, an aspect extraction algorithm is introduced, which automatically extracts aspect words from massive review texts and then classifies these extracted aspect words into different aspect categories. Secondly, nine novel aspect features are proposed for identifying spam reviews, which are classified into the review centric type and user centric type. The values of aspect features are calculated based on the results of the aspect extraction module. Finally, the XGBoost model is trained to classify spam reviews and non-spam reviews.

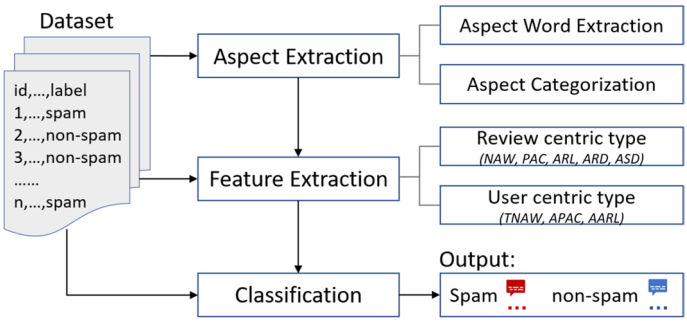

Fig. 1 The overall framework of ABCM

3.1 Aspect extraction

In comparison with reviewers, products, and reviews which can be easily identified in structured review datasets, it is a quite complicated and challenging task to extract product aspects from unstructured review texts. In this study, we propose an aspect extraction algorithm, including aspect word extraction and aspect categorization, to select the top n popular (wildly used) aspect words from the review texts and then classify the selected aspect words into k categories.

3.1.1 Aspect word extraction

In this study, we utilize the Bi-LSTM model to extract aspect words and related user opinions from the review texts. The aspect word extraction task is formulated as a sequence labeling task as follows. Given an input sequence $X = \{x_1, x_2, \cdots, x_n\}$, where x_i represents the ith word in the input sentence, we predict a sequence of labels $Y = \{y_1, y_2, \cdots, y_n\}$ for each word in the sentence, where y_i represents A (aspect word), E (opinion word), or O (other word). Opinion words usually are emotional words (such as 'good') that indicate a user's opinion on a certain product aspect (such as '*exterior*'). We automatically elicit the tuple of $<A$ (aspect word), E (opinion word)$>$ from the review texts using the Bi-LSTM model [13], which is a well-known state-of-the-art machine learning method with outstanding performance for sequence labeling.

The Bi-LSTM model includes an input sentence, an embedding layer, three Bi-LSTM layers, a softmax layer, and an output layer. Given an input sentence, the model predicts a label corresponding to each of the input tokens in the sentence. First, through the embedding layer, the input sentence is represented as a vector X. Then three Bi-LSTM layers are used to extract both the preceding and subsequent contextual information of each word. A softmax activation function on top of Bi-LSTM layers is used to calculate a probability distribution p over a set of predicates $\{A, E, O\}$. Finally, a list of labels for each word in the input sentence is predicted according to the corresponding output of the softmax layer.

For example, given an input sentence like '*Exterior is beautiful*', the model will output the labels of A, E, and O for the words '*exterior*', '*beautiful*', and '*is*', respectively. Then a tuple $<$*exterior*, *beautiful*$>$ can be obtained, in which '*exterior*' is an aspect word, and '*beautiful*' is the user's opinion on this aspect ('*exterior*'). The aspect word extraction module outputs two kinds of words: aspect words and opinion words, which will be used in the following steps.

3.1.2 Aspect categorization

As user-submitted review contents are generally not well structured, there are lots of flexible and variant aspect words contained in the review texts. To summarize the massive aspects and classify semantically similar aspects into the same category, we use the K-means clustering model for aspect categorization based on the similarity of these aspect words.

Given M extracted aspect words, we select the top n popular aspect words for aspect categorization, which forms an aspect word set $A = \{a_1, a_2, \cdots, a_n\}$. Firstly, each aspect word is represented by a vector, which is a word embedding learned from the review texts using the word2vec model. Then, with vector representations, the similarity between every two vectors (aspect words) is measured by the Euclidean distance. Finally, based on the semantic similarities, the m aspect words are classified into k aspect categories using the

K-means clustering model, i.e.,$AC = \{A_1, A_1, \cdots, A_i\}\left(\bigcap_{i=1}^{k} A_i = \varnothing, \bigcup_{i=1}^{k} A_i = A\right)$, in which each aspect category AC contains many aspect words from the set A, and each aspect word in the set A is classified into one specific aspect category.

3.2 Feature extraction

Since spammers may not have real experience in purchasing or using the product, the aspects depicted in spam reviews might be different from those in genuine reviews [37]. In this study, we propose nine novel aspect features which are classified into the review centric and user centric types, as listed in Table 1. Review centric aspect features refer to the micro aspect information contained in the reviews, and user centric aspect features refer to the macro information of aspects commented on by the users in their reviews.

To start describing the proposed aspect features, we list the terms and their notations that will be used in this paper as follows for ease of presentation:

- u, p, r: a user, a product, and a review, respectively;
- R_u: the set of reviews written by user u;
- R_p: the set of reviews on product p;
- P_u: the set of products reviewed by user u;
- ac: an aspect category;
- R_{ac}: the set of reviews on aspect category ac;
- AW_r: the set of aspect words in review r;
- AC_r: the set of aspect categories in review r.

3.2.1 Review centric aspect features

Number of Aspect Words (NAW): The number of aspect words in a review is a basic aspect feature. We define the *NAW* as the number of unique aspect words in a review r by Eq. (1).

$$NAW = \left|AW_r\right| \tag{1}$$

Percentage of Aspect Categories (PAC): In comparison with the aspect categories of a product that are discussed by all the users, an individual review often only involves a small number of aspect categories of the product. Then, we define the *PAC* of a review r on a product p as the number of aspect categories discussed in the review

Table 1 Proposed aspect features for review spam detection

Review centric type	User centric type
Number of Aspect Words (*NAW*)	Total Number of Aspect Words (*TNAW*)
Percentage of Aspects Categories (*PAC*)	Average Percentage of Aspects Categories (*APAC*)
Aspect Review Length (*ARL*)	Average Aspect Review Length (*AARL*)
Aspect Rating Deviation (*ARD*)	
Aspect Sentiment Deviation (*ASD*)	
Aspect Sentiment and Rating Deviation (ASRD)	

r divided by the total number of aspect categories of the product p, as shown in Eq. (2).

$$PAC = \frac{|AC_r|}{NA_p} \tag{2}$$

where NA_p denotes the total number of aspect categories discussed in all the reviews on the product p.

Aspect Review Length (ARL): A review is composed of a number of review sentences, of which some are related to particular aspects, while others are not. Then, the length of review sentences on particular aspects becomes valuable information to describe the characteristic of the aspects. We define the *ARL* of a review r by the length of sentences on the aspects divided by the length of the review r, as shown in Eq. (3):

$$ARL = \frac{\sum_{ac \in AC_r} L\left(st^r_{ac}\right)}{L(r)} \tag{3}$$

where st^r_{ac} represents the review sentences on aspect category ac in review r, the length L of review (sentences) is calculated by the number of words in the review (sentences).

Aspect Rating Deviation (ARD): We define the *ARD* as the difference between the numerical rating of a review and the sentiment score of the aspect categories discussed in the review. The sentiment score *SS* of an aspect category is calculated using an SVM model based on the opinion words on the aspect category, which are extracted by the Bi-LSTM model in the aspect extraction step. Given a piece of opinion words, the SVM model outputs a sentiment score of 1, 0, and 0.5, which represent positive, negative, and neutral classes, respectively. The calculation of the *ARD* is shown in Eq. (4):

$$ARD = Avg_{ac \in AC_r} \frac{\left|4SS\left(opw^r_{ac}\right) + 1 - Rating(r)\right|}{4} \tag{4}$$

where apw^r_{ac} represents the opinion words on aspect category ac in review r, $Rating(r)$ denotes the numerical rating of review r on a 5-scale, the sentiment score is also converted into a 5-scale by $4SS + 1$.

Aspect Sentiment Deviation (ASD): We define the *ASD* as the difference between the sentiment score of the discussed aspect categories in a review r of a product p and the average sentiment score of the discussed aspect categories in all the reviews on the product p, as shown in Eq. (5).

$$ASD = Avg_{ac \in AC_r} \left| SS\left(opw^r_{ac}\right) - Avg_{r' \in R_{ac}} SS\left(opw^{r'}_{ac}\right) \right| \tag{5}$$

Aspect Sentiment and Rating Deviation (ASRD): We define the *ASRD* as the difference between the average sentiment score of the discussed aspect categories in a review r of a product p and the average rating of the other reviews of the product, as shown in Eq. (6).

$$ASRD = \frac{\left| Avg_{ac \in AC_r}\left(4SS\left(opw^r_{ac}\right) + 1\right) - Avg_{r' \in R_p} Rating\left(r'\right) \right|}{4} \tag{6}$$

3.2.2 User centric aspect features

Total Number of Aspect Words (TNAW): At the reviewer/user level, we define the *TNAW* of a review r written by a user u as the total number of unique aspect words published by the user u, as shown in Eq. (7).

$$TNAW = \sum\nolimits_{r \in R_u} |AW_r| \tag{7}$$

Average Percentage of Aspect Categories (APAC): Corresponding to the *PAC* of a review r of a product p written by a user u, we define the *APAC* as the number of aspect categories discussed by the user u divided by the total number of aspect categories of the product p that the user u have commented on, as shown in Eq. (8):

$$APAC = \frac{\sum_{r \in R_u} |AC_r|}{\sum_{p \in P_u} NA_p} \tag{8}$$

where NA_p represents the total number of the aspect categories of the product p.

Average Aspect Review Length (AARL): Corresponding to the *ARL* of a review r written by a user u, we define the *AARL* as the average length of review sentences on all the aspect categories discussed by the user u, as shown in Eq. (9).

$$AARL = Avg_{r \in R_u} \frac{\sum_{ac \in AC_r} L\left(st^r_{ac}\right)}{L(r)} \tag{9}$$

3.3 Classification

We use the XGBoost (eXtreme Gradient Boosting) model to classify spam reviews and non-spam reviews, which is a well-designed gradient-boosted decision tree algorithm with state-of-the-art advantages in machine learning and data mining fields. XGBoost is an ensemble learning model, in which decisions from multiple machine learning models are combined to reduce errors and improve prediction when compared to a single machine learning model. In addition, the maximum voting technique is used on aggregated decisions to reduce the final prediction.

4 Experimental study

In this section, the performance of ABCM is empirically evaluated on two labeled Yelp datasets. Firstly, descriptions of the evaluation dataset are presented. Then the experimentation details are provided. At last, comparison experiments and evaluation metrics are designed.

4.1 Dataset description

The labeled YelpChi dataset [36] has been widely used in many previous studies and has been proven suitable and effective for supervised spam review detection with credible labels [36, 41, 46, 59]. There are two sub-datasets in the YelpChi dataset, i.e., the YelpChi_Hotel dataset for hotel reviews and the YelpChi_Res dataset for restaurant reviews.

Table 2 Basic information of the reviews in the Yelp datasets

Dataset	YelpChi_Hotel			YelpChi_Res		
	Spam	Non-spam	Total	Spam	Non-spam	Total
No. of reviews	779	4,897	5,676	8,301	56,969	65,270
No. of products	70	68	70	98	103	103
No. of users	750	4,148	4,898	7,116	27,541	34,519
No. of sentences	6,671	53,547	60,218	55,744	581,106	637,328
No. of unique words	9,285	26,254	28,424	31,787	99,541	108,505
No. of unique aspect words	2,121	4,321	6,385	13,147	35,083	47,080

After removing the reviews without information on products, users, and review content, the YelpChi_Hotel dataset contains 779 spam reviews and 4,897 non-spam reviews written by 4,898 users on 70 hotels in Chicago, and the YelpChi_Res dataset includes 8,301 spam reviews and 56,969 non-spam reviews written by 34,519 users on 103 restaurants in Chicago (see Table 2). The data items of the datasets used in this paper are date, review_id, user_id, product_id, rating, review_content, and label (spam or non-spam). Examples of genuine reviews and spam reviews in the two datasets are listed as follows. These examples reveal that spammers have difficulty in describing the detailed experiences of purchasing or using products, thus their comments on the fine-grained product attributes are short and straightforward, such as "Loved the location" and "Great Restaurant".

- YelpChi_Hotel dataset: (Genuine) *This hotel is great. Location, room, concierge, gym, everything. By the way, they do not replenish the so-famous bathroom products anymore. Instead, they leave some other branded products. (Spam) Loved it! Loved the location! Loved the staff! Loved the beds! Loved the showers! Just flat out loved it! Even the food in the restaurant on the main floor!*
- YelpChi_Res dataset: (Genuine) *Food was very fresh and tasty. Great service and free valet parking are bonus. The portions are reasonably sized. Prices are decent. (Spam) Great Restaurant, Great Price, Great Food. Try the Chicken Vesuvio delivered. Incredible and mouth watering!!!!!*

4.2 Experimental settings

4.2.1 Aspect extraction settings

We first cut all review sentences in the Yelp datasets into sub-sentences according to the punctuations like ",.!?;", then randomly selected and manually tagged 400 hotel reviews (3,703 sentences) from the YelpChi_Hotel dataset and 400 restaurant reviews (3,779 sentences) from the YelpChi_Res datasets, respectively, to train two separated Bi-LSTM models for aspect word extraction. For each Bi-LSTM model, the proportions of these selected review sentences for training, validation, and test sets were 70%, 15%, and 15%, respectively. The detailed parameter settings of the BiLSTM model after tuning are listed in Table 3 (a).

The parameters of the aspect extraction settings are the number of aspect words n and the number of aspect categories k. There are no commonly accepted rules to set the parameters appropriately, these parameters are dependent on the dataset of the task. For the two

Table 3 Parameter settings of the BiLSTM model and the XGBoost model

(a) BiLSTM model		(b) XGBoost model	
Parameter	Value	Parameter	Value
learning_rate	0.01	n_estimators	500
hidden_units	100	learning_rate	0.01
epochs	200	max_depth	10
batch_size	32	min_child_weight	1
input_dim	5,562	subsample	0.95
output_dim	300	colsample_bytree	0.95
input_length	140	reg_alpha	0.01
dropout_rate	0.2	gamma	0.4
optimizer	'Adadelta'		

Yelp datasets in this study, we totally extracted $|A|$= 6,385 unique hotel aspect words and $|A|$= 47,080 unique restaurant aspect words (Table 2) through the BiLSTM models. However, a large number of the extracted aspect words are meaningless or unrelated to the product aspects. To guarantee that the aspect words are representative and are contained in the majority of reviews, yet to exclude infrequent aspect words, we determine the optimal number of the most popular aspect words to be used for review spam detection, n, by setting n increase from 1 to $|A|$ until p (the proportion of the reviews which contain at least one of the top n popular aspect words) equals to P_A (the proportion of the reviews which contain at least one of the aspect words in A), in which A is the set of all the extracted aspect words, and $|A|$ is the number of aspect words in A. Taking the YelpChi_Hotel dataset for example, Algorithm 1 illustrates the process of determining the optimal number of the most popular hotel aspect words for review spam detection.

After obtaining the optimal numbers of the aspect words, we then classify them into k categories using the K-means clustering model. The number of aspect categories k is set as {2, 4, 6, 8, 10, 12, 14} for obtaining the most reasonable results.

Algorithm 1 Determining the optimal number of the most popular aspect words for review spam detection

Inputs: a set of 6,385 hotel aspect words A and a set of 5,676 hotel reviews R

Outputs: the optimal number of most popular hotel aspect words n

$p_A \leftarrow$ 99.07% (the percentage of the reviews which contain at least one of the aspect words)
$A \leftarrow$ order(A) // Rank the aspect words by their occurrences
$R_A \leftarrow \varnothing$
for $n = 1, 2, \ldots, |A|$ **do**
  for $i = 1, 2, \ldots, |R|$ **do**
    if $a_n \in r_i$ **then** // $a_n \in A,\ r_i \in R$
      $R_A \leftarrow R_A \cup \{ r_i \}$
      break
    end if
  end for
  $p \leftarrow |R_A| / |R|$ // Calculate the proportion of reviews
  if $p \geq p_A$ **then** // n is enough
    break
  end if
end for
return n

4.2.2 Classification model settings

Noting that the numbers of spam and non-spam reviews in the datasets are imbalanced, we employ under-sampling to randomly select a sub dataset from the majority class (non-spam reviews) according to the number of samples in the minority class (spam reviews). Then the selected sub dataset is combined with the minority class to form a balanced class distribution data for model training.

The parameters of the XGBoost model for classification are tuned by the five-fold cross-validation method. That is, we split the training set into 5 folds randomly, in which 4 folds are used to train the model and the remaining onefold is used for validation. We carefully tune the parameters by a grid search with a small but adaptive step size. For example, the learning rate is tuned from {0.0001, 0.0001, 0.001, 0.01, 0.1}. After several trials, we obtain the optimal values of the parameters, as shown in Table 3(b). Then the classification models are trained and tested over two labeled Yelp datasets and evaluated using five-fold cross-validation, this process is repeated five times to get an average performance.

4.3 Comparison experiments

To evaluate the effectiveness of the proposed aspect features for review spam detection, we select nine textual features and six behavior features that were widely used in previous studies [20, 25, 29, 33, 36] for comparison. The nine textural features include review length (RL), percentage of words with all letters capitalized (PCW), percentage of capital letters (PC), ratio of 1st person pronouns (RFPP), ratio of exclamation sentences containing '!' (RES), ratio of subjective words (SW), ratio of objective words (OW), length of unigrams-based description (DLU), and length of bigrams-based description (DLB). The six behavior features include repeated review (RR), extreme rating (ER), early time review (ETR), is singleton review (ISR), rating deviation (RD), and review sentiment deviation (RSD).

Then we develop XGBoost models with different combinations of the three feature groups, i.e., textual features (T), behavior features (B), aspect features (A), textual and behavior features (T + B), textual and aspect features (T + A), behavior and aspect features (B + A), and all features (B + T + A). We also compare the performance of ABCM with other baseline models which are widely used in the review spam detection field, including SVM (Support Vector Machine), LR (Logistic Regression), RF (Random Forest), CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), and LSTM (Long Short-Term Memory). For comparison, the inputs of these models are the nine proposed aspect features, all models are repeated five times and evaluated using the five-fold cross validation method.

4.4 Evaluation metrics

We choose Accuracy (A), Recall (R), Precision (P), F1 score (F), and area under a ROC curve (AUC) as evaluation metrics which are widely adopted in related areas [46].

5 Results and discussions

In this section, we first present the results of aspect extraction, and then compare the performance of ABCM with other baseline and state-of-the-art approaches. In addition, we discuss the computational complexity of ABCM and evaluate the effect of aspect settings (i.e.,

number of aspect words and number of aspect categories) on model performance. Finally, we highlight the theoretical and practical implications of the proposed aspect features for identifying spam reviews.

5.1 Aspect extraction results

In the aspect extraction step, the Bi-LSTM models achieved precisions of 0.84 and 0.92 for the YelpChi_Hotel and YelpChi_Res datasets, respectively. After removing duplicated words and stop words, we finally extracted 6,385 unique hotel aspect words and 47,080 unique restaurant aspect words. The numbers of occurrences of these aspect words in the datasets are shown in Fig. 2(a) and (b), revealing that the majority of aspect words were discussed with low frequency, and only several common aspect words were wildly discussed. For example, the aspect word "room" has been discussed 9,671 times in 66.9% (3,800) of hotel reviews by 60% (3,404) of reviewers, and "food" has been discussed 57,286 times in 44.3% (28,882) of restaurant reviews by 29.3% (19,110) of reviewers.

It is worth noting that the extracted 6,385 hotel aspect words were contained in 99.07% of the hotel reviews, and the extracted 47,080 restaurant aspect words were contained in 98.73% of the restaurant reviews. Also, there are many product-unrelated words in the extracted aspect words. Based on the algorithm for determining the optimal number of aspect words for review spam detection (Algorithm 1), we obtained $n = 644$ most popular hotel aspect words and $n = 898$ most popular restaurant aspect words after deliberate experiments. The distributions of the occurrences of the selected aspect words are shown in Fig. 2(b) and (e), indicating that most reviews contain a small number of aspect words, and each review usually contains 5 to 6 unique aspect words in both datasets. For example,

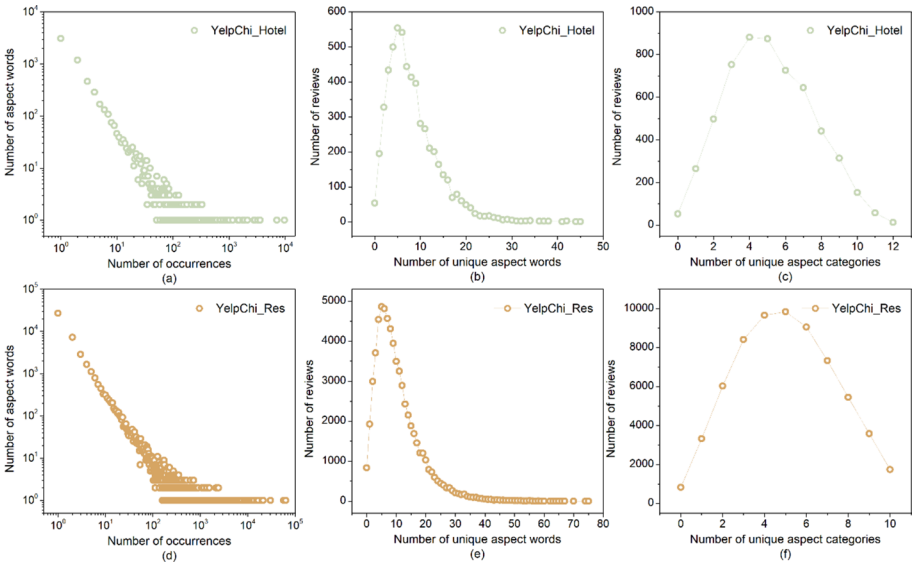

Fig. 2 Distributions of the occurrences of all the extracted (**a**) 6,385 hotel aspect words and (**b**) 47,080 restaurant aspect words. Distributions of the numbers of the selected (**c**) 644 hotel aspect words and (**d**) 898 restaurant aspect words in each review. Distributions of the numbers of the (**e**) 12 hotel aspect categories and (**f**) 10 restaurant aspect categories in each review

90.1% of hotel reviews contain less than 15 unique aspect words and 90.6% of restaurant reviews contain less than 20 unique aspect words.

Finally, the selected 644 hotel aspect words and 898 restaurant aspect words are classified into $k = 12$ categories and $k = 10$ categories, respectively. Among a few values we tried for the cluster number, they generated the most reasonable results on our data. The distributions of the number of aspect categories are shown in Fig. 2(c) and (f), indicating that most of the reviews only discussed 3 to 6 aspect categories. About 85.7% (4,865) of hotel reviews discussed less than 8 aspect categories, and 87.3% (57,006) of restaurant reviews discussed less than 9 aspect categories. Averagely, each hotel (restaurant) review contains 5.5 (5.8) aspect categories. In addition, we present the aspect words in different aspect categories in Fig. 3. We can see that the aspect words are clustered into meaningful aspect categories based on the semantic relationships between aspect words. For example, the hotel aspect words in category #9 are related to the aspect category "service" and the restaurant aspect words in category #3 are related to the aspect category "employee".

The above results indicate that users usually discuss particular product attributes in their reviews, while in each review, the number of discussed aspects is quite limited compared with the total number of words. And it is common that each review usually contains 4 to 6 aspects of the product, while a part of reviews contains even fewer aspect words. This suggests the potential of aspect-related clues for distinguishing spam reviews and non-spam reviews.

5.2 Classification results and comparison

We first evaluate the proposed aspect features compared with the widely used textual features and behavior features in review spam detection based on the YelpChi_Hotel dataset (hotel domain) and YelpChi_res dataset (restaurant domain). Table 4 shows that aspect features combined with behavior features achieved the highest performance across both hotel and restaurant domains. While textual features and behavior features both only yielded less than 70% precision, aspect features achieved 96.5% and 85.1% precision in the YelpChi_Hotel and YelpChi_Res datasets, respectively. With the use of aspect features, the classification performance compared with textual features or/and behavior features (i.e., T, B,

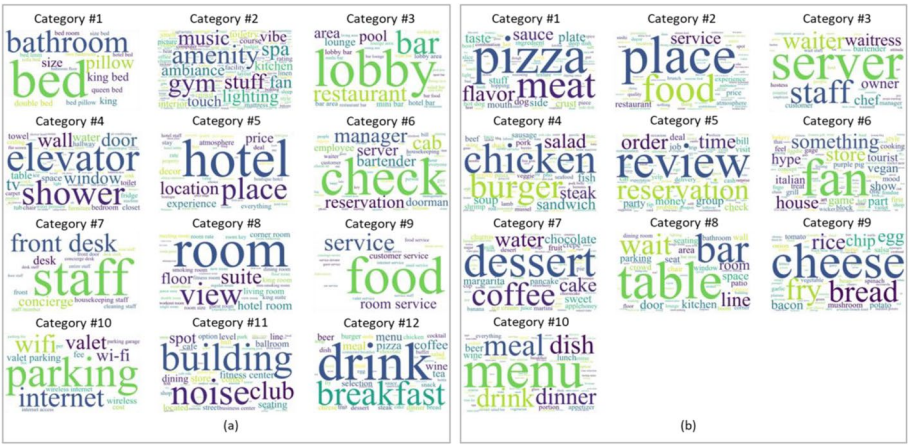

Fig. 3 Hotel aspect words (**a**) and restaurant aspect words (**b**) in different categories

Table 4 Classification results of XGBoost model with different features on Yelp datasets

Dataset	Features	Accuracy	Recall	Precision	F1 score	AUC
YelpChi_Hotel	T	56.8	56.8	56.8	56.7	59.3
	B	67.7	67.7	68.0	67.6	77.3
	T + B	70.5	70.5	70.6	70.5	80.3
	A	**96.4**	**96.4**	**96.5**	**96.4**	**99.0**
	T + A	96.3	96.3	96.4	96.3	98.9
	B + A	97.7	97.7	97.7	97.7	99.3
	T + B + A	97.5	97.5	97.6	97.5	99.3
YelpChi_Res	T	61.2	61.2	61.2	61.2	65.3
	B	65.8	65.8	65.8	65.8	72.2
	T + B	68.5	68.5	68.7	68.4	74.7
	A	**84.5**	**84.5**	**85.1**	**84.5**	**93.5**
	T + A	83.9	83.9	84.6	83.9	93.2
	B + A	85.4	85.4	85.6	85.4	94.4
	T + B + A	85.2	85.2	85.4	85.1	94.2

* T: textual, B: behavior, A: aspect; All the improvements of aspect features over textual or/and behavior features are statistically significant with $p < 0.0001$ based on paired t-test
The bold entries represent our proposed method (i.e., XGBoost based on aspect features)

and T + B) has been improved by 18.44 ~ 38.86% in the hotel domain and 16.11 ~ 28.01% in the restaurant domain, respectively. All the improvements are statistically significant at the confidence level of 95% based on paired t-test ($p < 0.0001$).

Regarding the XGBoost classification model used in our method, we then compare it with the baseline machine learning models such as SVM, LR, RF, CNN, RNN, and LSTM. Table 5 shows that the proposed aspect features outperformed behavior and textual features in identifying spam reviews for all the above models, and the XGBoost model with aspect features achieved the best AUC across both hotel and restaurant domains, which significantly outperformed baseline models with $p < 0.01$ based on paired t-test. In general, XGBoost obtains better performance than single machine learning models such as SVM, LR, and RF, but the difference in AUC between XGBoost and other deep learning models such as CNN is relatively small.

Table 5 AUC of XGBoost and baseline models with different features on Yelp datasets

Dataset	Features	SVM	LR	RF	CNN	RNN	LSTM	XGBoost
YelpChi_Hotel	T	59.9	60.2	57.7	61.0	59.6	60.0	59.3
	B	77.0	78.7	75.3	78.9	69.3	69.4	77.3
	T + B	80.3	80.8	77.7	80.9	70.8	71.4	70.5
	A	**98.6**	**98.5**	**98.2**	**98.9**	**96.4**	**96.6**	**99.0**
YelpChi_Res	T	65.9	65.9	60.5	65.9	59.6	63.0	65.3
	B	63.1	65.8	71.3	66.8	61.1	58.3	72.2
	T + B	73.3	73.1	69.5	74.4	66.7	69.2	68.4
	A	**87.0**	**86.7**	**89.6**	**91.1**	**82.5**	**80.9**	**93.5**

* T: textual, B: behavior, A: aspect
The bold entries signify the best performance of the models

Finally, we compare the proposed ABCM with several methods in previous studies on the same datasets and find that ABCM also shows competitive performance in identifying spam reviews, as shown in Table 6. For example, Mukherjee et al. [36] used bigrams and behavior features that achieved the highest precision of 86.7% and 84.1% in the YelpChi_Hotel and YelpChi_Res datasets, respectively. When combining YelpChi_Hotel and YelpChi_Res as the YelpChi dataset, ABCM also achieved an AUC of 93.9%, which is higher than the results on the same dataset using other methods, such as the SpEagle (unsupervised model, AUC = 78.87%) [41], the SPR2EP (semi-supervised model, AUC = 80.71%) [57], and the HFAN (supervised model, AUC = 83.24%) [59].

5.3 Complexity analysis

In the aspect extraction step, the BiLSTM model is used to extract aspect words and user opinion words from massive review content, with computational complexity $O(n \times d^2)$, where d denotes the dimension of representation and n is the length of the sequence. In the classification step, the computational complexity of the XGBoost model used for identifying spam reviews based on aspect features is $O(m \times n \times log(n) + k \times d \times m \times n)$, where m represents the number of features, n is the number of samples in the dataset, d stands for the number of estimators in the XGBoost model, and d denotes the depth of each tree.

In summary, the computational complexity of the BiLSTM model keeps the same as the LSTM architecture, since only an extra LSTM cell is added to learn backward. The XGBoost model used in this study retains the same computational complexity as the orginal XGBoost model.

5.4 Effect of aspect settings

Based on the selected 644 hotel aspect words and 898 restaurant aspect words, we evaluate the impact of the aspect settings on the performance of ABCM. Figure 4 and Fig. 5 show the precisions of ABCM under different proportions of aspect words (p) and different numbers of aspect categories (k) in each dataset. In the YelpChi_Hotel dataset, ABCM achieves the best precision when we classify the 644 hotel aspect words (p=1) into k = 14

Table 6 Comparison of techniques based on the Yelp datasets

Dataset	Technique	Model	Features	Accuracy	Recall	Precision	F1 score	AUC
YelpChi_Hotel	Mukherjee et al. [36]	SVM	Bi + B	84.8	82.5	86.7	84.5	/
	ABCM	XGBoost	A	96.4	96.4	96.5	96.4	99.0
YelpChi_Res	Mukherjee et al. [36]	SVM	Bi + B	86.1	87.3	84.1	85.7	/
	ABCM	XGBoost	A	84.5	84.5	85.1	84.5	93.5
YelpChi	SpEagle [41]	LBP	T + B + R	/	/	/	/	78.87
	SPR2EP [57]	LR	T + R	/	/	/	/	80.71
	HFAN [59]	TransH	S	/	/	/	/	83.24
	ABCM	XGBoost	A	85.0	85.0	85.5	84.9	94.0

* LBP: Loopy Belief Propagation, TransH: Knowledge Graph Embedding by Translating on Hyperplanes, Bi: bigrams, B: behavior, A: aspect, T: textual, R (relational), S (semantic)

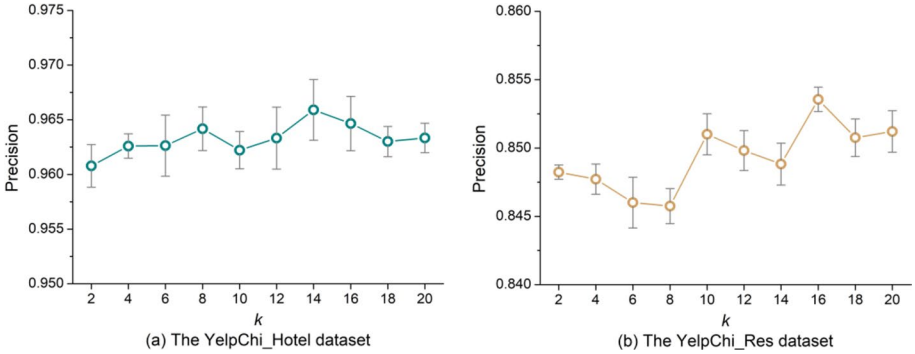

Fig. 4 Performance of ABCM with different configurations of the number of aspect categories k based on (**a**) 644 hotel aspect words and (**b**) 898 restaurant aspect words

categories, as shown in Fig. 4(a) and Fig. 5(a). In the YelpChi_Res dataset, ABCM optimizes its performance among all the configurations when we classify the 898 restaurant aspect words ($p=1$) into $k=16$ categories, as shown in Fig. 4(b) and Fig. 5(b).

As observed from Fig. 4 and Fig. 5, we can see that ABCM achieved lower precision under small numbers of aspect categories, and the precision of ABCM improved slightly with the increase of the number of aspect words. A large number of aspect words could help improve the performance in identifying spam reviews. However, the difference in model performance among different aspect settings is very small (about 1%), which indicates the robustness of ABCM. Therefore, ABCM is capable of successfully identifying spam reviews using a small number of aspect words, which can not only reduce the computation costs but also can be easily extended for spam detection in different product reviews.

5.5 Implications of aspect features

This study highlights the importance of aspect-related features in review spam detection, which carries several theoretical implications as follows. Compared with previous studies which used just one aspect-related feature for identifying spam reviews [54, 58], our study proposed a list of novel aspect features which show more outstanding performance in

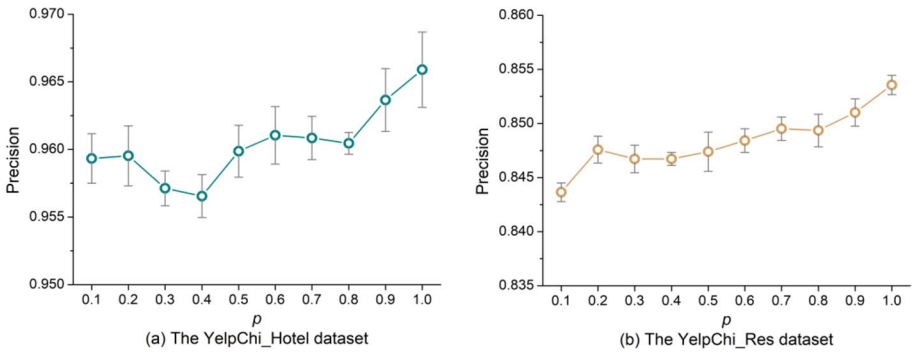

Fig. 5 Performance of ABCM with different configurations of the proportion of aspect words p. The selected hotel and restaurant aspect words are classified into (**a**) 12 categories and (**b**) 10 categories, respectively

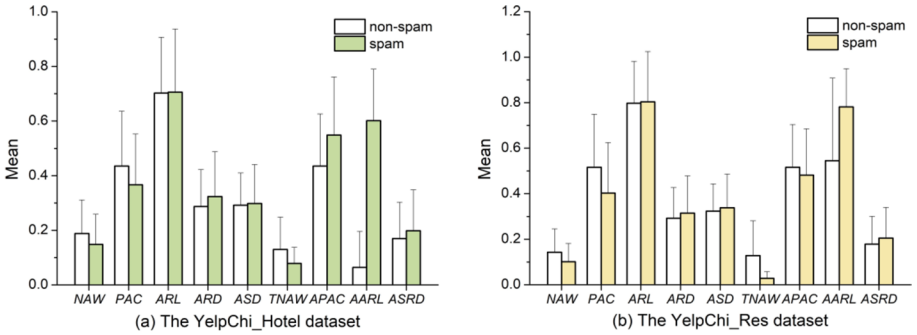

Fig. 6 Mean of the aspect features for spam reviews and non-spam reviews in (**a**) the YelpChi_Hotel dataset and (**b**) the YelpChi_Res dataset

review spam detection than the widely used textual and behavior features. In previous studies, the aspects have been mainly extracted by topic models such as LDA (Latent Dirichlet Allocation), and thus the number of aspects extracted and used for review spam detection is relatively small (less than 100) [54, 58]. To address this problem, our study proposed an aspect extraction algorithm using the Bi-LSTM model and the K-means clustering model, which automatically extracts a large number of valuable aspects (about 1,000). In addition, this study reveals that the textual features are not good indicators for review veracity (see Table 5), which is consistent with the findings in previous studies [3, 36].

The findings of this study also carry several practical implications. Two-sample t-test (independent t-test) results show that spam reviews have significantly smaller NAW (number of aspect words) and PAC (percentage of aspect categories) but higher ARD (aspect rating deviation) and ASRD (aspect sentiment and rating deviation) than non-spam reviews in both datasets, as shown in Fig. 6. It suggests that a review with fewer aspect words or aspect categories is more likely to be a spam review, and a review can be fake if there is a large difference between the rating of the review (or the average rating of the product reviews) and the sentiment score of the discussed aspects in this review. From the user perspective, it is found that the users in spam reviews show significantly lower TNAW (total number of aspect words) but higher AARL (average aspect review length) than the users in non-spam reviews. That is, spammers have commented on a large number of product aspects in spam reviews to make their experience look truthful, however, due to the lack of true experience with the products, their review content on particular product aspects is relatively short. The above findings provide valuable information for consumers and manufacturers to identify spam reviews. Also, the findings on the user centric aspect features could also help identify spammers.

5.6 Evaluation on other platforms

To evaluate the performance of the proposed method on other online review platforms, we apply ABCM on a labeled real-world Amazon dataset [18]. The data items of these Amazon reviews include _id, asin (i.e., productId), category, class (spam or non-spam), helpful, overall (i.e., rating), reviewText, reviewTime, reviewerId, reviewerName, summary, and unixReviewTime. Firstly, we randomly selected 132 hotel-related products' reviews (called Amazon_HK_hotel dataset) and 210 restaurant-related products' reviews (called Amazon_HK_res

Table 8 Classification results of XGBoost model with different features on Amazon datasets

Dataset	Features	Accuracy	Recall	Precision	F1 score	AUC
Amazon_HK_hotel	T	59.9	59.9	60.1	59.8	64.1
	B	66.4	66.4	67.0	66.1	73.3
	T + B	69.5	69.5	70.1	69.3	78.2
	A	**82.2**	**82.2**	**82.5**	**82.2**	**90.6**
Amazon_HK_res	T	61.4	61.4	61.5	61.4	65.2
	B	70.2	70.2	70.8	70.0	76.3
	T + B	72.4	72.4	73.1	72.2	80.5
	A	**80.7**	**80.7**	**81.0**	**80.6**	**88.4**

* T: textual, B: behavior, A: aspect. All the improvements of aspect features over textual or/and behavior features are statistically significant with $p < 0.0001$ based on paired t-test

The bold entries signify the best performance of the models

dataset) from the "Home_and_Kitchen" category.[1] The Amazon_HK_hotel dataset contains 2,105 spam reviews and 6,772 non-spam reviews, and the Amazon_HK_res dataset includes 2,250 spam reviews and 11,061 non-spam reviews. Secondly, we extracted aspect words and opinion words from the review content by the BiLSTM model, and classified the most popular aspect words into different aspect categories using the K-means algorithm. In detail, the most popular 1,005 aspect words in the Amazon_HK_hotel dataset and the most popular 1,121 aspect words in the Amazon_HK_res dataset are clustered into 16 categories and 20 categories, respectively. Then, we calculated the proposed aspect features of the labeled reviews, and calculated the behavior features and textual features as introduced in Sect. 4.3 for comparison. It is worth noting that the ratings of all the spam reviews and non-spam reviews in the Amazon dataset are {1, 2, 3} and {4, 5}, respectively. That is, we can fully correctly identify spam reviews from the Amazon dataset by only screening the review ratings. Therefore, when calculating behavior features, we removed the RD (rating deviation) feature which is calculated based on the ratings of reviews. Finally, we performed the XGBoost model on two balanced datasets which are constructed according to Sect. 4.2.2.

Evaluation results of the XGBoost model with different features are shown in Table 8. It is obvious that the proposed aspect features also outperformed the behavior and textual features in the Amazon datasets, which achieved 82.5% and 81.0% precision in the Amazon_HK_hotel and Amazon_HK_res datasets, respectively. With the use of aspect features, the classification performance compared with textual features or/and behavior features (i.e., T, B, and T + B) has been improved by 12.03 ~ 26.24% in the Amazon_HK_hotel dataset and improved by 7.75 ~ 22.88% in the Amazon_HK_res dataset. All the improvements are statistically significant at the confidence level of 95% based on paired t-test ($p < 0.0001$).

6 Conclusion

This paper proposes an aspect-based classification method called ABCM for review spam detection using a list of novel aspect features. The Bi-LSTM model is used to automatically extract massive aspect words from the review texts, and the K-means clustering model is

[1] https://www.kaggle.com/datasets/naveedhn/amazon-product-review-spam-and-non-spam?select=Home_and_Kitchen.

used to classify the aspect words into several aspect categories. Based on the aspect extraction results, we propose nine novel aspect features to capture fine-grained spamming clues regarding product attributes, and then train the XGBoost model to classify the reviews as spam and non-spam.

The use of the proposed aspect features has not been studied before in the review spam detection field. Interestingly, this set of features has brought great results reaching a maximum of 96.5% precision on the YelpChi_Hotel dataset, and outperforms the widely used textual and behavior features by about 16.11 ~ 38.86%. The proposed ABCM outperforms the baseline models and state-of-the-art approaches used in previous studies. In addition, ABCM shows outstanding performance in identifying spam reviews from both Yelp.com and Amazon.com. This fact indicates that the idea of extracting aspect features for review spam detection has been successful, as the detailed information of aspects discussed in spam reviews and non-spam reviews may show different characteristics.

However, the current study only evaluates the review spam detection model on two kinds of products (i.e., hotels and restaurants). In the future, it is promising to apply ABCM on different kinds of products and services provided by other platforms, such as TripAdvisor.com and Resellarratings.com, to evaluate the effectiveness and generalizability of adapting aspect features in machine learning methods for online review quality control.

Data availability The Yelp datasets used in the current study are available at http://liu.cs.uic.edu/download/yelp_filter/, and the Amazon datasets used in this study are available at https://www.kaggle.com/datasets/naveedhn/amazon-product-review-spam-and-non-spam?select=Home_and_Kitchen.

Declarations

Conflict of interest This work was supported by the National Natural Science Foundation of China (72201272, 72025405, 72088101), the National Social Science Foundation of China (22ZDA102), the Hunan Science and Technology Plan Project (2020TP1013, 2020JJ4673, 2023JJ40685), the Shenzhen Basic Research Project for Development of Science and Technology (JCYJ20200109141218676, 202008291726500001), the Innovation Team Project of Colleges in Guangdong Province (2020KCXTD040), and the Social Science Foundation of Hunan Province (20YBA012). The authors declare that they have no conflict of interest.

References

1. Akoglu L, Chandy R, Faloutsos C (2013). Opinion fraud detection in online reviews by network effects. Seventh Int AAAI Conf Weblogs and Social Media, 7(1), pp.2–11. https://ojs.aaai.org/index.php/ICWSM/article/view/14380. Accessed 2013-07-10
2. Bajaj S, Garg N, Singh S (2017) A novel user-based spam review detection. Procedia Computer Science 122:1009–1015. https://doi.org/10.1016/j.procs.2017.11.467
3. Barbado R, Araque O, Iglesias AC (2019) A framework for fake review detection in online consumer electronics retailers. Inf Process Manage 56(4):1234–1244. https://doi.org/10.1016/j.ipm.2019.03.002
4. Bhuvaneshwari P, Rao AN, Robinson YH (2021) Spam review detection using self attention based CNN and bi-directional LSTM. Multimedia Tools and Applications 80:18107–18124. https://doi.org/10.1007/s11042-021-10602-y
5. Buettner R (2016) Predicting user behavior in electronic markets based on personality-mining in large online social networks: A personality-based product recommender framework. Electron Mark 27:247–265. https://doi.org/10.1007/s12525-016-0228-z
6. Cai M, Tan Y, Ge B, Dou Y, Huang G, Du Y (2021) PURA: A product-and-user oriented approach for requirement analysis from online reviews. IEEE Syst J 99:1–12. https://doi.org/10.1109/JSYST.2021.3067334
7. Chua AY, Banerjee S (2016) Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. Comput Hum Behav 54:547–554. https://doi.org/10.1016/j.chb.2015.08.057

8. Dong M, Yao L, Wang X, Benatallah B, Huang C, Ning X (2018) Opinion fraud detection via neural autoencoder decision forest. Pattern Recogn Lett 132:21–29. https://doi.org/10.1016/j.patrec.2018.07.013
9. Etaiwi W, Naymat G (2017) The impact of applying different preprocessing steps on review spam detection. Procedia Computer Science 113:273–279. https://doi.org/10.1016/j.procs.2017.08.368
10. Fei G, Mukherjee A, Liu B, Hsu M, Castellanos M, Ghosh R (2013). Exploiting burstiness in reviews for review spammer detection. Seventh Int Conf Weblogs and Social Media, pp.175–184
11. Feng S, Banerjee R, Yejin C (2012). Syntactic stylometry for deception detection. 50th Annual Meet Assoc Comput Linguist, pp.171–175
12. Gao Y, Gong M, Xie Y, Qin QK (2020) An attention-based unsupervised adversarial model for movie review spam detection. IEEE Trans Multimedia 23:784–796. https://doi.org/10.1109/TMM.2020.2990085
13. Graves A, Mohamed AR, Hinton G (2013). Speech recognition with deep recurrent neural networks. Int Conf Acoustics, Speech, and Signal Process (ICASSP), pp.6645–6649. https://doi.org/10.1109/ICASSP.2013.6638947
14. He D, Pan M, Hong K, Cheng Y, Chan S, Liu X, Guizani N (2020) Fake review detection based on PU learning and behavior density. IEEE Network 99:1–6. https://doi.org/10.1109/MNET.001.1900542
15. Hernández Fusilier D, Montes-y-Gómez M, Rosso P, Guzmán Cabrera R (2015) Detecting positive and negative deceptive opinions using PU-learning. Inf Process Manage 51(4):433–443. https://doi.org/10.1016/j.ipm.2014.11.001
16. Hernández-Castañeda Á, Calvo H, Gelbukh A, Flores J (2017) Cross-domain deception detection using support vector networks. Soft Comput 21(3):585–595. https://doi.org/10.1007/s00500-016-2409-2
17. Heydari A, Tavakoli M, Salim N, Heydari Z (2015) Detection of review spam: A survey. Expert Syst Appl 42(7):3634–4364. https://doi.org/10.1016/j.eswa.2014.12.029
18. Hussain N, Mirza H, Hussain I, Iqbal F, Memon I (2020) Spam review detection using the linguistic and spammer behavioral methods. IEEE Access 8:53801–53816. https://doi.org/10.1109/ACCESS.2020.2979226
19. Jia S, Zhang X, Wang X, Liu Y (2018). Fake reviews detection based on LDA. 4th Int Conf Inf Manag, pp.280–283. https://doi.org/10.1109/INFOMAN.2018.8392850
20. Jindal N, Liu B (2008). Opinion spam and analysis. Int Conf Web Search and Data Mining, pp. 219–230. https://doi.org/10.1145/1341531.1341560
21. Jindal N, Liu B (2007). Review spam detection. 16th international conference on World Wide Web, pp.1189–1190. https://doi.org/10.1145/1242572.1242759
22. Karami A, Zhou B (2015). Online review spam detection by new linguistic features. Proceedings of iConference 2015. http://hdl.handle.net/2142/73749. Accessed 2015-03-15
23. KC S, Mukherjee A (2016). On the temporal dynamics of opinion spamming: Case studies on Yelp. 25th Int Conf World Wide Web, pp.369–379. https://doi.org/10.1145/2872427.2883087
24. Li H, Fei G, Wang S, Liu B, Shao W, Mukherjee A, Shao J (2017). Bimodal distribution and co-bursting in review spam detection. 26th Int Conf World Wide Web, pp.1063–1072. https://doi.org/10.1145/3038912.3052582
25. Li F, Huang M, Yang Y, Zhu X (2011). Learning to identify review spam. IJCAI Proc-Int Joint Conf Artificial Intell pp.2488–2493. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-414
26. Li H, Liu B, Mukherje A, Shao J (2014) Spotting fake reviews using positive-unlabeled learning. Computación y Sistemas 18(3):467–475. https://doi.org/10.13053/cys-18-3-2035
27. Li J, Ott M, Cardie C, Hovy E (2014). Towards a general rule for identifying deceptive opinion spam. 52nd Annual Meet Assoc Comput Linguist, pp.1566–1576. https://doi.org/10.3115/v1/P14-1147
28. Li A, Qin Z, Liu R, Yang Y, Li D (2019). Spam review detection with graph convolutional networks. 28th ACM Int Conf, pp.2703–2711. https://doi.org/10.1145/3357384.3357820
29. Lim E, Nguyen V, Jindal N, Liu B, Lauw H (2010). Detecting product review spammers using rating behaviors. 19th ACM Int Conf Inf Knowledge Manag, pp.939–948. https://doi.org/10.1145/1871437.1871557
30. Lu Y, Zhang L, Xiao Y, Li Y (2013). Simultaneously detecting fake reviews and review spammers using factor graph model. Third Annual ACM Web Science Conference, pp.225–233. https://doi.org/10.1145/2464464.2464470
31. Luo Y, Tang R (2019) Understanding hidden dimensions in textual reviews on Airbnb: An application of modified latent aspect rating analysis (LARA). Int J Hosp Manag 80:144–154. https://doi.org/10.1016/j.ijhm.2019.02.008
32. Mukherjee S, Dutta S, Weikum G (2016). Credible review detection with limited information using consistency features. European Conf Machine Learning and Principles and Practice of Knowledge Discovery, pp.195–213. https://doi.org/10.1007/978-3-319-46227-1_13

33. Mukherjee A, Kumar A, Liu B, Wang J, Hsu M, Castellanos M, Ghosh R (2013). Spotting opinion spammers using behavioral footprints. 19th ACM SIGKDD Int Conf Knowledge Discovery and Data Mining, pp.632–640. https://doi.org/10.1145/2487575.2487580
34. Mukherjee A, Liu B, Glance N (2012). Spotting fake reviewer groups in consumer reviews. 21st Annual Conf World Wide Web, pp.191–200. https://doi.org/10.1145/2187836.2187863
35. Mukherjee A, Liu B, Wang J, Glance N, Jindal N (2011). Detecting group review spam. 20th Int Conf Companion on World Wide Web, pp.93–94. https://doi.org/10.1145/1963192.1963240
36. Mukherjee A, Venkataraman V, Liu B, Glance N (2013). What yelp fake review filter might be doing?. Seventh Int Conf Weblogs and Social Media, pp.409–418
37. Noekhah S, Salim N, Zakaria NH (2018). A comprehensive study on opinion mining features and their applications. International conference of reliable information and communication technology. Int Conf Reliable Inf Commun Technol, pp.78–89. https://doi.org/10.1007/978-3-319-59427-9
38. Noekhah S, Salim N, Zakaria N (2019) Opinion spam detection: Using multi-iterative graph-based model. Inf. Process. Manage 57(1):102140. https://doi.org/10.1016/j.ipm.2019.102140
39. Ott M, Cardie C, Hancock J (2013). Negative deceptive opinion spam. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.497–501
40. Rastogi A, Mehrotra M (2017) Opinion spam detection in online reviews. J Inf Knowl Manag 16(4):1750036. https://doi.org/10.1142/S0219649217500368
41. Rayana S, Akoglu L (2015). Collective opinion spam detection: Bridging review networks and metadata. 21th ACM SIGKDD Int Conf Knowledge Discovery and Data Mining, pp.985–994. https://doi.org/10.1145/2783258.2783370
42. Ren Y, Ji D (2017) Neural networks for deceptive opinion spam detection: An empirical study. Inf Sci 385(38):213–224. https://doi.org/10.1016/j.ins.2017.01.015
43. Shahariar GM, Biswas S, Omar F, Shah F, Hassan S (2019). Spam review detection using deep learning. IEEE 10th Annual Inf Technol, Electronics and Mobile Commun Conf, pp.0027–0033. https://doi.org/10.1109/IEMCON.2019.8936148
44. Shehnepoor S, Salehi M, Farahbakhsh R, Crespi N (2017) NetSpam: A network-based spam detection framework for reviews in online social media. IEEE Trans Inf Forensics Secur 12(7):1585–1595. https://doi.org/10.1109/TIFS.2017.2675361
45. Shojaee S, Murad M, Azman A, Sharef N, Nadali S (2013). Detecting deceptive reviews using lexical and syntactic features. Int Conf Intell Syst Des Appl, pp.53–58. https://doi.org/10.1109/ISDA.2013.6920707
46. Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsl 19(1):22–36. https://doi.org/10.1145/3137597.3137600
47. Tang X, Qian T, You Z (2020) Generating behavior features for cold-start spam review detection with adversarial learning. Inf Sci 526:274–288. https://doi.org/10.1016/j.ins.2020.03.063
48. Thapa R, Lamichhane B, Ma D, Jiao X (2021). SpamHD: Memory-efficient text spam detection using brain-inspired hyperdimensional computing. IEEE Comput Soc Annual Symposium on VLSI (ISVLSI), pp.84–89. https://doi.org/10.1109/ISVLSI51109.2021.00026
49. Tsai CF, Chen K, Hu YH, Chen WK (2020) Improving text summarization of online hotel reviews with review helpfulness and sentiment. Tour. Manag 80:104122. https://doi.org/10.1016/j.tourman.2020.104122
50. Wang X, Liu K, Zhao J (2017). Handling cold-start problem in review spam detection by jointly embedding texts and behaviors. 55th Annual Meet Associ Comput Linguist pp.366–376. https://doi.org/10.18653/v1/P17-1034
51. Wang H, Lu Y, Zhai C (2010). Latent aspect rating analysis on review text data: A rating regression approach. ACM SIGKDD Int Conf Knowledge Discovery and Data Mining, pp.783–792. https://doi.org/10.1145/1835804.1835903
52. Wang G, Xie S, Liu B, Yu P (2012) Identify online store review spammers via social review graph. ACM Trans Intell Syst Technol 3(4):1–21. https://doi.org/10.1145/2337542.2337546
53. Xie S, Wang G, Lin S, Yu P (2012). Review spam detection via temporal pattern discovery. ACM SIGKDD Int Conf Knowledge Discovery and Data Mining, pp.823–831. https://doi.org/10.1145/2339530.2339662
54. Xue H, Wang Q, Luo B, Seo H, Li F (2019) Content-aware trust propagation toward online review spam detection. Journal of Data and Information Quality 11(3):1–31. https://doi.org/10.1145/3305258
55. Yang Y, Mueller N, Croes R (2016) Market accessibility and hotel prices in the Caribbean: The moderating effect of quality-signaling factors. Tour Manage 56:40–51. https://doi.org/10.1016/j.tourman.2016.03.021

56. Ye J, Akoglu L (2015). Discovering Opinion Spammer Groups by Network Footprints. ACM on Conf Online Social Netw, pp.97. https://doi.org/10.1145/2817946.2820606
57. Yilmaz C, Durahim O (2018). SPR2EP: A semi-supervised spam review detection framework. IEEE/ACM Int Conf Advances in Social Networks Analysis and Mining, pp.306–313. https://doi.org/10.1109/ASONAM.2018.8508314
58. You L, Peng Q, Xiong Z, He D, Qiu M, Zhang X (2019) Integrating aspect analysis and local outlier factor for intelligent review spam detection. Futur Gener Comput Syst 102:163–172. https://doi.org/10.1016/j.future.2019.07.044
59. Yuan C, Zhou W, Ma Q, Lv S, Han J, Hu S (2019). Learning review representations from user and product level information for spam detection. IEEE Int Conf Data Mining, pp.1444–1449. https://doi.org/10.1109/ICDM.2019.00188
60. Zhang W, Du Y, Yoshida T, Wang Q (2018) DRI-RCNN: An approach to deceptive review identification using recurrent convolutional neural network. Inf Process Manage 54(4):576–592. https://doi.org/10.1016/j.ipm.2018.03.007
61. Zhang M, Fan B, Zhang N, Wang W, Fan W (2021) Mining product innovation ideas from online reviews. Inf Process Manag 58:102389. https://doi.org/10.1016/j.ipm.2020.102389