Neural Networks Merging Semantic and Non-semantic Features for Opinion Spam Detection

Chengzhi Jiang and Xianguo Zhang*

College of Computer Science, Inner Mongolia University, Hohhot 010021, China
chengzhi@mail.imu.edu.cn; 2595083628@qq.com

Abstract. In recent years, abundant online reviews on products and services have been generated by individuals. Since customers may refer to relevant online reviews when shopping, the existence of fake reviews can affect potential consumption. Opinion spam detection has attracted widespread attention from both the business and research communities. In this paper, a neural network model combining the semantic and non-semantic features based on the detailed feature exploration is established to detect opinion spams. First, the model learns discourse feature representation with hierarchical attention neural networks which can capture local and global semantic information. And then we synthesis the non-semantic features with multi-kernel convolution neural networks. Finally, the last state vectors of the two-feature learning networks are concatenated and taken as input to the softmax layer for classification. Experiments show that the proposed model is very effective and we get 0.853 AUC which outperforms the baseline methods. Besides, the experiment results on an additional dataset also indicate robustness of this identification model.

Keywords: Opinion spam, Deceptive review, Semantic features, Non-semantic features, Neural networks, Hierarchical attention mechanism.

1 Introduction

With the popularity of e-commerce and online review websites, an increasing number of online consumers are well adapted at sharing and exchanging their feelings and opinions by posting reviews on the web. Online reviews play a major role in consumers' decisions, and research has shown that consumers' purchase decisions and sales are significantly affected by user-generated online reviews of products and services [3]. However, the valuable and informative reviews give businesses strong motivations to manipulate their reputations on the Internet. By posting fake reviews, such malicious individuals and groups are involved in promoting their targeted products or services, or defame certain competitors. Jindal and Liu [9] defined such individuals as opinion spammers, whose activities were called opinion spamming. Deceptive opinion spam is a more insidious type of opinion spam with fictitious opinions which are deliberately written to sound authentic [23]. It is difficult for consumers to directly discern whether a review is deceptive. These deceptive reviews are likely to mislead potential

consumers and have attracted significant attention from both business and research communities.

The objective of opinion spam detection is to identify whether the review is true or fake, so it can be considered a binary-category classification problem. Most of existing methods on opinion spam detection are training a classifier with various discrete features extracted from the labeled dataset. Classical classification features, e.g. POS, emotional polarity and n-gram, can represent linguistic and emotional information. Previous researches show effective features enable to give strong performance for identification [20, 23]. Hence, it's necessary to develop forceful feature engineering. Neural networks have been widely used in natural language processing tasks, due to the advantage of capturing local and global semantic feature, and have achieved good performance recently [13].

Since a piece of review is commonly short document, which has the hierarchical structure: words make up sentences, and sentences constitute documents. In this work, we build a hierarchical attention network (HAN) to capture reviews' semantic features and detect deceptive reviews, and this network has two main stages. In the first stage, a long short-term memory (LSTM) is used to produce sentence presentation from word presentation. Then employing the bidirectional gated recurrent neural network (GRNN) to learn a review presentation from the sentence presentation in the second stage. Besides, the feed-forward networks with attention are added into both layers of the model to capture more important lexical and syntax information. The review representation learned by hierarchical model can be used as classification features to detect deceptive opinion spam.

In addition to semantic features, some special features from metadata of reviews, reviewers and businesses have been explored. In this work, these features containing little semantic and lexical information are defined as non-semantic features. Rather than traditional discrete feature presentation, we build a matrix of feature sequences and regard this feature matrix as input of neural network with multiple convolution kernels to synthesis non-semantic features effectively. As a result, a novel neural network model merging semantic and non-semantic features (MFNN) is proposed for opinion spam detection. Results on development experiments show that MFNN significantly outperforms the state-of-the-art detection models.

The several major contributions of the work presented in this paper are as following:

- We present a HAN model to learn document-level presentation. Compared with a single neural network structure, hierarchical neural network is easier to learn continuous representations of reviews.
- We explore a set of non-semantic features for opinion spam detection, which is represented by feature embedding method. Such feature representations trained by convolution neural networks improves the recognition ability of model.
- We verify the performance of MFNN in different domains, and experiments show that our model has the generalization ability.

2 Related Work

2.1 Deceptive Opinion Spam Detection

With the ever-increasing popularity of the Internet, a variety of spams have brought plenty of troubles to general people. In the past years, spam detection research mainly focuses on web spam and E-mail spam [2, 4, 21]. Usually, web spam and E-mail spam have obvious characteristics, such as irrelevant keywords or URLs. But the clues to fake reviews are subtle. Jindal and Liu [8] first began to study opinion spam problem. According to analyze the Amazon product reviews, they presented three main types of spam reviews and proposed several classification techniques to distinguish them.

Machine learning technology is the mainstream research method for spam detection. Yoo and Gretzel [34] gathered a small amount of hotel reviews and analyzed their linguistic difference. By employing Amazon Mechanical Turkers to write fake reviews, Ott et al. [23] created a gold standard dataset and improved the classification performance with LIWC. Since then, a line of subsequent works based on this benchmark dataset [5, 22] have been presented. Various of nonmachine learning techniques to opinion spam detection have also been explored, such as pattern matching [9, 14, 37] and graph-based methods [30, 31]. However, that only can be applied in certain types of review spamming activities.

Existing works have exploited features outside the review content itself as well. For example, Li et al. [14] built a robust identification model with n-gram features as well as POS and LIWC features on their cross-domain datasets. Mukherjee et al. [19] used the data crawled from the Yelp.com to extract a few users' behavioral features and they proved the effectiveness of behavioral features.

2.2 Neural Networks for Representation Learning

The field of natural language processing is an important application area for deep learning. Xu and Rudniky [32] first proposed the idea of using neural networks to train language models. Recently, distribute word representation has been used by quantity of models for representation learning. For example, Mikolov et al. [17, 18] proposed two word embedding structures of CBOW and Skip-gram and tried to improve the calculation speed of the model using negative sampling and hierarchical softmax. Pennington et al. [24] utilized Glove, the embedding model of global word-word co-occurrence, to import word embedding.

As for the presentation learning of sentences and documents, numerous methods have been proposed. Mikolov et al. [18] introduced paragraph vector to learn document presentations. Socher et al. [27] proposed learning sentence-level semantic composition from recursive neural networks. Hill et al. [7] proposed learning distributed presentation of sentences from unlabeled data. Considering the capture of n-gram information, convolution neural networks have been widely used for presentation learning [10, 36]. Researchers have proposed various of recurrent neural network models for learn the document semantic [6, 15, 29].

Table 1. Dataset statistics

Items	Values
Domain	Restaurant
Fake	8261
Truthful	58631
Total #_reviews	66892
#_reviewers	34962
#_businesses	129

Attention mechanism model refers to a current neural network with an attention mechanism [25], and it is suitable for a variety of tasks such as computer vision and natural language processing. Because the attention mechanism can capture latent and important features from training data, Yang et al. [33] proposed the hierarchical attention networks for document classification.

3 Data and Feature Sets

3.1 Data Set

In this work, we choose to use real-life authentic labeled reviews filtered from Yelp.com [19]. Yelp is a well-known large-scale online review website and its filtering algorithm can filter some fake or suspicious reviews. Although Yelp's fake review filtering is not perfect, it's a commercial review hosting site that has been preforming industrial scale filtering [28]. The dataset is unbalanced clearly from Table 1. Although data imbalance may affect the performance of classification model, the fake reviews in real life are really minority class. Thus, we conduct the experiments with the full dataset ignoring the problem of data imbalance.

3.2 Features Exploration and Analysis

In this paper, the two features, i.e. semantic features and non-semantic features, are as inputs to opinion spam detection model. The former is the knowledge learned from the text of the review, which is used to describe the meanings of words and sentences. The latter is mainly extracted from reviews, reviewers and businesses itself. Next, we introduce the process of exploration and analysis of features.

The neural network models apply the look-up matrix layer to map the words into corresponding word embeddings which are low dimensional, continuous and real-valued vectors. In this work, we have pre-trained word embeddings of 123,152 and 100 dimensions using the continuous bag-of-words model architecture on the open Yelp dataset[1]. During training, the words out of vocabulary are initialized randomly.

Many previous studies have proved that the customers' behavioral characteristics have a significant influence on the identification performance of deceptive opinion

[1] https://www.yelp.com/dataset/

spams [11, 35, 19], thus we intuitively extract the following features from the metadata as non-semantic features. The Cumulative Distribution Function (CDF) is plotted to analyze the difference between spam and non-spam among these features as Fig.1.

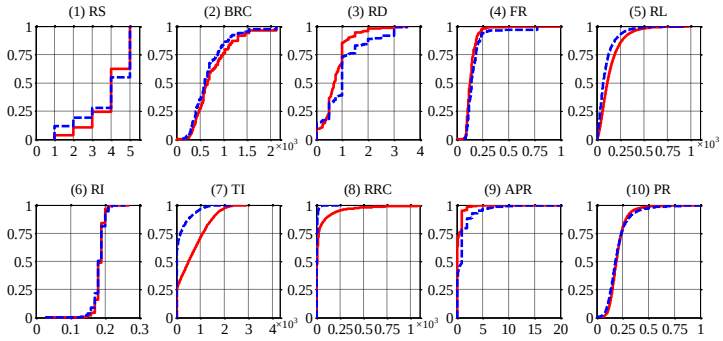

Fig. 1. CDF of non-semantic features. Cumulative percentage of non-spam (in red/solid) and spam (in blue/dotted) vs. non-semantic feature value

Rating Score (RS): The higher the rating score, the more positive the reviewer is towards the business. It can be found that the spam review is more likely to give a low rating in Fig.1(1).

Business Review Count (BRC): From Fig.1(2), the proportion of fake reviews in all business reviews is slightly higher than true reviews. It may be the clue that merchants need a lot of reviews to expand the discussion of their goods or services, and these reviews may be fake.

Rating Deviation (RD): To measure the rating deviation of a reviewer, the absolute score bias of a rating score on business from business's rating is computed. Then, we calculate the average score bias of a reviewer on its all reviews. From the Fig.1(3), we can find that the value of about 85% of true reviews is less than 1. And 10% of the spammers have the deviation of not less than 3.

Filtering Ratio (FR): Our intuition is that if most reviews of a business are filtered by Yelp's filter, a newly posted review on this business is more likely to be fake. It can be found from Fig.1(4) that about 5% of the businesses associated with spam reviews have the filtering ratio 0.25-0.75.

Review Length (RL): Spammers are often hired and required to complete a certain number of spam reviews, so they generally don't spend a lot of time writing reviews. As shown in Fig.1(5), a majority of spam reviews have shorter length than real reviews.

Readability Index (RI): The readability of review's content may affect the customer's feelings when they read that review. Researchers have evaluated readability of online reviews by ARI and CLI [12], which are related to the number of characters, words and sentences per reviews. The performance is well using ARI and CLI as features for opinion spam detection. Experiments show that the ratio of ARI and CLI also have ability to make a contribution to the result of detection, and the performance is better than ARI and CLI. Thus, the ratio is used to denote the readability index.

$$ARI = 4.71\times(\frac{characters}{words})+0.5\times(\frac{words}{sentences})-21.43 \quad (1)$$

$$CLI = 5.89\times(\frac{characters}{words})-0.3\times(\frac{sentences}{words})-15.8 \quad (2)$$

Time Interval (TI): We show the CDF of maximal time interval of all reviews posted by same reviewer in Fig.1(7). More than half of spammers have very small time intervals, and 55% of spammers posted all reviews by a time gap less than 10. That might mean that these spammers have discarded their accounts after posting a spam review before too long.

Reviewer Review Count (RRC): This feature refers to the number of reviews that a reviewer has. From Fig.1(8), about 90% of the spammers have post fewer than 13 posts, but 30% of non-spammers have posted more than 30 posts.

Average Posting Rate (APR): The activity level of reviewers can be measured by this metric. Fig.1(9) shows the posting frequency of 95% of real reviewers is less than 2, and more than 10% of spammers have a posting rate which is greater than 2. This is related to the fact that spammers need to post a certain amount of deceptive reviews.

Punctuation Ratio (PR): When people write reviews, they probably add some special symbols to express their feelings, such as ":)". Meanwhile, ones often use a series of exclamation marks or question marks to express strong emotions. We take these factors into account and the ratio of punctuation marks to the review length is used to indicate this situation.

Labeled-LDA (LLDA): Latent Dirichlet Allocation (LDA) is a type of topic generation model and it is an important text modeling model in the field of text mining and information processing, which can extract latent topics from text data [1, 16, 26]. In our work, Labeled-LDA features are trained by Stanford Topic Modeling Toolbox[2] and the label distribution of each word is gotten. Based on the appearance times of words under each label, the most relevant top M words to each label are selected. Relevant words of all labels are merged as LLDA feature words, and the frequency $P_{i,j}$ of word w_i under each label j is as feature value $L(w_i)$. We combine w_i and $L(w_i)$ into a dict $\{w_1: L(w_1), w_2: L(w_2) \ldots w_{2M}: L(w_{2M})\}$ as our LLDA feature.

The Pearson correlation analysis is used to evaluate the non-semantic features excepting LLDA feature. The correlation coefficients are shown in Table 2. It is generally argued that features are considered highly relevant if coefficients are greater than 0.5. As shown in the table, the features are basically irrelevant, which means the latent information they contain is not duplicated. Therefore, we apply these features to our identification model.

4 Methodology

In this section, we present the details of our proposed MFNN model, which can learn

[2] https://nlp.stanford.edu/software/tmt/tmt-0.4

Table 2. Pearson correlation coefficients of non-semantic features

	RS	BRC	RD	FR	RL	RI	TI	RRC	APR	PR
RS	1.000									
BRC	0.082	1.000								
RD	-0.435	-0.064	1.000							
FR	0.022	-0.391	0.024	1.000						
RL	-0.114	0.014	0.034	-0.021	1.000					
RI	-0.102	0.076	0.029	-0.033	0.088	1.000				
TI	-0.017	-0.075	-0.100	-0.016	0.125	0.018	1.000			
RRC	-0.003	-0.029	-0.062	-0.015	0.116	-0.004	0.439	1.000		
APR	0.016	0.014	0.036	0.008	-0.092	-0.013	-0.322	-0.017	1.000	
PR	0.074	0.011	-0.012	-0.013	-0.197	-0.422	-0.003	-0.010	0.012	1.000

discourse representation of documents and synthesize information of non-semantic features. The model mainly consists of two parts: document-level modeling and non-semantic feature modeling as shown in Fig.2.

4.1 The Modeling Process of All Features

In Section 3.2, we have mentioned that the feature values of LLDA are not one dimensional and they denote the frequency of feature words appearing under a label. Thus, as shown in the lower right of Fig.2, the word's LLDA feature and its word vector are combined as the new word representation.

The document generally is with a hierarchical structure: words make up sentences, and sentences constitute documents. Thus, we build the hierarchical networks to capture documents' semantic features. The structure is shown in the left of Fig.2. Firstly, sentence representation is learned from word embeddings, and then the document representation is generated from sentence vector. Finally, since some important words or sentences in the document can promote performance, the feed-forward network with attention layer [25] is respectively added to each representation learning layer to capture this information effectively. It is worth mentioning that both LSTM and GRNN can capture information in text sequences very well, but experiments show that the neural networks in Fig.2 can achieve better results.

In our experiments, the non-semantic features of all one-dimensional discrete values are presented as the feature dict. Each key r_{id} of the dict is the sequence number of the mapped review, and the value corresponding to each key is a set of feature sequences. The feature dict is as $\{r_{id}: (RS_{id}, BRC_{id} \dots PR_{id})\}$. Thus, all feature sequences can compose a feature matrix $R^{D\times L}$, where D is the number of reviews and L is the length of feature sequences. This feature matrix is used as the input of the non-semantic feature learning model. The embedding layer is applied to distribute the uniform and random weight values of the fully connected layer on non-semantic features. To capture the local information, the multi-kernel convolutional neural networks is utilized to synthesize non-semantic features with width of 3, 4, and 5. As a result, that the learned feature vector combines with the document vector is input into the softmax layer for classification.

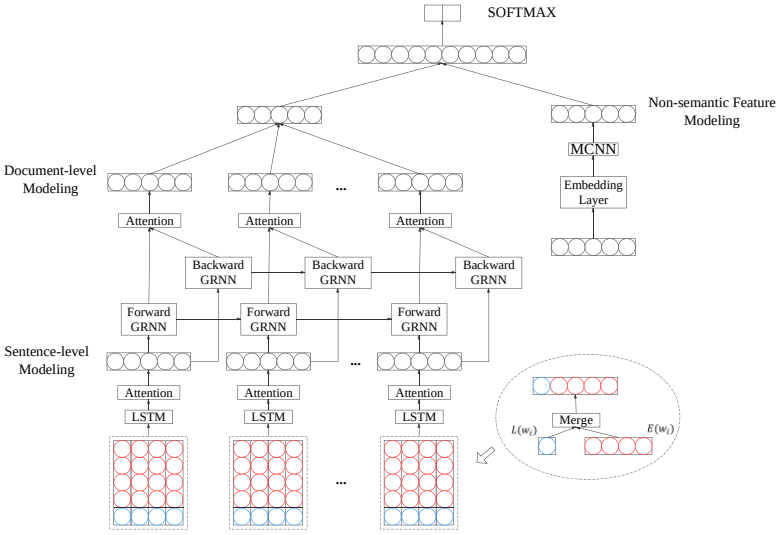

Fig. 2. Neural networks merging semantic and non-semantic Features

4.2 Evaluation Metrics

Accuracy is the most common evaluation metric for classification. However, in the case of binary classification and unbalanced dataset, especially when we are more interested in the minority class, employing accuracy to evaluate model performance is not appropriate. Thus, we choose the metric of Area Under the Curve (AUC) to evaluate performance of detection model. The Receiver Operating Characteristic (ROC) curve is a comprehensive indicator based on sensitivity and specificity drawn by different thresholds with TPR and FTR as the coordinate axes. AUC is the area under the ROC curve. The larger the area under the curve or the closer the curve is to the upper left corner (TPR=1, FPR=0), the better the model.

5 Experiments and Validation

We use all features mentioned above to verify the performance of our proposed opinion spam detection model. In our experiments, we utilize 80% of the samples as the training set, 10% as the validation set, and 10% as the test set. The validation set is used to optimize the hyper-parameters of neural networks.

5.1 Development Experiments and Validation

We choose to use respectively unigram feature and bigram feature to conduct our baseline based on SVM with 5-fold cross-validation, which was done in [19, 23]. To compare different classification models, we conducted a set of development experiments. The all classification models are as Table 3.

Table 4 shows the results of all development experiments on restaurant field. From

Table 3. Description of all experimental models

Features	Model	Description
Semantic Only	Unigram	Using word unigram feature in SVM with 5-fold cross validation
	Bigram	Using word bigram feature in SVM with 5-fold cross validation.
	Average	Simply using the average of all word vectors as the review vector.
	CNN_1	A multi-kernel CNN is used, and the last state vector of neural network is used as the document vector.
	RNN	A single-directional RNN is used and its last state vector of neural network is used as the review vector.
	BLSTM	A bidirectional LSTM is used and its last state vector of neural network is used as the review vector.
	HAN	Hierarchical neural network based on attention mechanism is used.
Non-semantic Only	SVM	Using discrete non-semantic features in SVM with 5-fold cross validation.
	CNN_2	A multi-kernel CNN is used, and the last state vector of neural network is used as the feature vector.
All	MFNN	The neural network model merging semantic and non-semantic features of this paper.

the table, the performance of recognition model using n-gram feature on the Yelp dataset is poor. However, previous experiments have shown that the classification performance based gold standard review dataset using n-gram feature can reach the better evaluation scores [23]. The reason may be that the gold dataset is collected by crowdsourcing websites and Turkers post reviews according to rules of the task. These fake opinion reviews are quite standard, but the real-world reviews from Yelp.com are noisy. And the data marked by the Yelp filter is not completely correct. Meanwhile, the performance of models using neural network structures is better than traditional machine learning methods according to the results. The AUC value of RNN model is the worst of several neural network models based on semantic features and we consider the reason is that RNN does not process long sequences efficiently resulting in gradient dispersion. From the AUC values of SVM and CNN_2 models, we can find that the non-semantic features have a significant promotion of recognition performance. Although non-semantic features can effectively facilitate opinion spam recognition, the information of review text is also very important. The experiment also proves that our syncretic model achieves the best AUC, which is 0.853.

To further validate the model performance, we apply an additional dataset of hotel field with 780 spam reviews and 5078 true reviews [19] on MFNN model. We repeat the same feature engineering to train hotel data. The results are largely consistent with those of restaurant data, which show that the MFNN model of this paper achieves the best classification performance AUC=0.923 as Table 4.

Table 4. The AUC values of all experiments

Model	Domain	
	Restaurant	Hotel
Unigram	0.496	0.545
Bigram	0.517	0.529
Average	0.639	0.631
CNN_1	0.713	0.667
RNN	0.599	0.538
BLSTM	0.726	0.672
HAN	0.731	0.751
SVM	0.786	0.858
CNN_2	0.800	0.885
MFNN	**0.853**	**0.923**

5.2 Experiment Extension

We experimentally study the effect of distribution of positive and negative samples in our proposed model. The ratios of spam to non-spam reviews in our two datasets are both up to 7:1, so we conduct a set of extension experiments by tuning the number of true reviews in the experimental data. The new experimental datasets are generated by random negative sampling techniques, in which the ratios of true and fake reviews are 1:1, 2:1, 3:1, 4:1, 5:1 and 6:1. The results is shown in Fig.3, from which we can see the AUC slightly fluctuates around 0.85 on the restaurant data, and it denotes that the distribution of samples has little effect. This may indicate that our model has some application significance in real life, after all, fake reviews are rare. In the hotel data, the AUC value fluctuates greatly. Considering the small amount of data in the dataset of hotel domain, and the maximum AUC value is obtained in the natural distributed dataset, but it does not prove that the more true reviews, the better the detection performance of fake reviews.

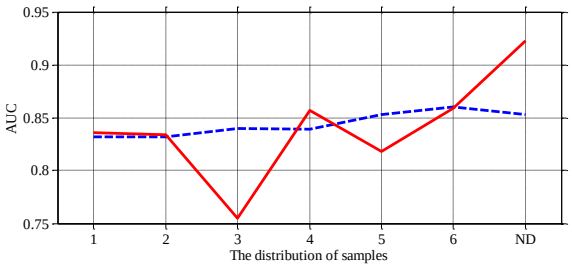

Fig. 3. The AUC of different sample distribution of restaurant (in blue/dotted) and hotel (in red/solid) areas

6 Conclusion

We introduce a novel neural network model merging semantic and non-semantic features for opinion spam detection. The experiment results show that the hierarchical neural network based on attention mechanism is better than the simple network. Non-semantic features have greatly promoted the performance of fake review detection. Through our work, we have explored a set of non-semantic features and employed a multi-kernel convolution neural network to synthesis these features. And the results show our proposed detection model outperforms the baseline method. Besides, the validation experiment also indicates that our model has better robustness.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022 (2003)
2. Castillo, C., Donato, D., Gionis, A., Murdock, V., Silvestri, F.: Know your neighbors: Web spam detection using the web topology. In: SIGIR 2007. pp. 423–430. ACM (2007)
3. Cui, G., Lui, H.K., Guo, X.: The effect of online consumer reviews on new product sales. International Journal of Electronic Commerce 17(1), 39–58 (2012)
4. Drucker, H., Wu, D., Vapnik, V.N.: Support vector machines for spam categorization. Trans. Neur. Netw. 10(5), 1048–1054 (1999)
5. Feng, S., Banerjee, R., Choi, Y.: Syntactic stylometry for deception detection. In: ACL 2012. pp. 171–175 (2012)
6. Glorot, X., Bordes, A., Bengio, Y.: Domain adaptation for large-scale sentiment classification: A deep learning approach. In: ICML 2011. pp. 513–520 (2011)
7. Hill, F., Cho, K., Korhonen, A.: Learning distributed representations of sentences from unlabelled data. pp. 1367–1377 (2016)
8. Jindal, N., Liu, B.: Analyzing and detecting review spam. In: ICDMW 2007. pp. 547–552 (2007)
9. Jindal, N., Liu, B.: Opinion spam and analysis. In: WSDM 2008. pp. 219–230 (2008)
10. Johnson, R., Zhang, T.: Effective use of word order for text categorization with convolutional neural networks (2014)
11. Ko, M.C., Huang, H.H., Chen, H.H.: Paid review and paid writer detection. pp. 637–645 (2017)
12. Korfiatis, N., GarcíA-Bariocanal, E., SáNchez-Alonso, S.: Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. Electron. Commer. Rec. Appl. 11(3), 205–217 (2012)
13. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: ICML 2014 (2014)
14. Li, H., Chen, Z., Liu, B., Wei, X., Shao, J.: Spotting fake reviews via collective positive-unlabeled learning. In: ICDM 2014. pp. 899–904. IEEE Computer Society (2014)
15. Li, L., Qin, B., Ren, W., Liu, T.: Document representation and feature combination for deceptive spam review detection. Neurocomputing 254 (2017)
16. Li, W.B., Sun, L., Zhang, D.K.: Text classification based on labeled-lda model. Chinese Journal of Computers 31, 620–627 (2009)
17. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. Computer Science (2013)

18. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. In: NIPS 2013. pp. 3111–3119 (2013)
19. Mukherjee, A., Venkataraman, V., Liu, B., Glance, N.: What yelp fake review filter might be doing? In: ICWSM 2013 (2013)
20. Mukherjee, A., venkataraman, V., Liu, B., Glance, N.: Fake review detection: Classification and analysis of real pseudo review. UIC-CS-03-2013 (2013)
21. Ntoulas, A., Najork, M., Manasse, M., Fetterly, D.: Detecting spam web pages through content analysis. In: WWW 2006. pp. 83–92 (2006)
22. Ott, M., Cardie, C., Hancock, J.: Estimating the prevalence of deception in online review communities. In: WWW 2012. pp. 201–210. ACM (2012)
23. Ott, M., Choi, Y., Cardie, C., Hancock, J.T.: Finding deceptive opinion spam by any stretch of the imagination. In: ACL 2011. pp. 309–319 (2011)
24. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. vol. 14, pp. 1532–1543 (2014)
25. Raffel, C., P. W. Ellis, D.: Feed-forward networks with attention can solve some long-term memory problems (2015)
26. Ramage, D., Hall, D., Nallapati, R., Manning, C.D.: Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In: EMNLP 2009. pp. 248–256 (2009)
27. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C., Ng, A., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. EMNLP1631, 1631–1642 (2013)
28. Stoppelman, J.: Why yelp has a review filter? (2009)
29. Tang, D., Qin, B., Liu, T.: Document modeling with gated recurrent neural network for sentiment classification. pp. 1422–1432 (2015)
30. Wang, G., Xie, S., Liu, B., Yu, P.S.: Review graph based online store review spammer detection. In: ICDM 2011. pp. 1242–1247 (2011)
31. Wang, G., Xie, S., Liu, B., Yu, P.S.: Identify online store review spammers via social review graph. ACM Trans. Intell. Syst. Technol. 3(4), 61:1–61:21 (2012)
32. Xu, W., Rudnicky, A.: Can artificial neural networks learn language models? In: ICSLP 2000 (2000)
33. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical attention networks for document classification. pp. 1480–1489 (2016)
34. Yoo, K.H., Gretzel, U.: Comparison of deceptive and truthful travel reviews. In: Information and Communication Technologies in Tourism 2009. pp. 37–47 (2009)
35. Zhang, D., Zhou, L., Luo Kehoe, J., Kilic, I.D.: What online reviewer behaviors really matter? effects of verbal and nonverbal behaviors on detection of fake online reviews. Journal of Management Information Systems 33, 456–481 (2016)
36. Zhang, X., Zhao, J., Lecun, Y.: Character-level convolutional networks for text classification (2015)
37. Zhou, L., Sung, Y.w., Zhang, D.: Deception performance in online group negotiation and decision making: The effects of deception experience and deception skill. Group Decision and Negotiation 22(1), 153–172 (2013)