Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2016, Article ID 4935792, 7 pages
http://dx.doi.org/10.1155/2016/4935792

Research Article

Exploiting Product Related Review Features for Fake Review Detection

Chengai Sun,[1,2] Qiaolin Du,[1] and Gang Tian[1]

[1]*College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China*
[2]*Wenshang County Zhihui Ziyou Information Technology Co., Ltd., Jining 272500, China*

Correspondence should be addressed to Gang Tian; tiangang@whu.edu.cn

Received 14 March 2016; Revised 27 June 2016; Accepted 4 July 2016

Academic Editor: Chunlin Chen

Product reviews are now widely used by individuals for making their decisions. However, due to the purpose of profit, reviewers game the system by posting fake reviews for promoting or demoting the target products. In the past few years, fake review detection has attracted significant attention from both the industrial organizations and academic communities. However, the issue remains to be a challenging problem due to lacking of labelling materials for supervised learning and evaluation. Current works made many attempts to address this problem from the angles of reviewer and review. However, there has been little discussion about the product related review features which is the main focus of our method. This paper proposes a novel convolutional neural network model to integrate the product related review features through a product word composition model. To reduce overfitting and high variance, a bagging model is introduced to bag the neural network model with two efficient classifiers. Experiments on the real-life Amazon review dataset demonstrate the effectiveness of the proposed approach.

1. Introduction

It has become more and more common for one to read online reviews before he/she make purchase decisions [1]. This gives high incentives for opinion spammers to write fake reviews to promote or to demote some target products or business. According to [2, 3], there are 2–6% fake reviews in Orbitz, Priceline, Expedia, Tripadvisor, and so forth. Mukherjee also reported that Yelp has a fake review rate of 14–20% [3]. Thus, detecting fake online reviews is becoming an important issue to ensure that the online reviews continue to be trusted materials of opinions, rather than being swarming with lies.

Researchers have proposed various fake review detection approaches in the past few years to preserve the accuracy of online opinion mining results. One major task in this area is to distinguish between fake reviews and truthful reviews [4]. A variety of methods were proposed to address this task mainly from two angles: reviewer and review. For example, the works in [4–6] mainly use content features of reviews to represent the reviews for classification tasks. On the other hand, the methods in [7–10] try to exploit the behaviour information of the reviewers to benefit the prediction task. Different from these works, we will examine the effects of product related review features for fake review detection.

Since when the spammers write the fake reviews, they tend to describe a product using some special feature words and sentimental words. It is helpful for the fake review detection model to capture these product related review features. Inspired by this, we proposed a convolutional neural network (CNN) model which captures the product related review features by a linear composition of products and reviews, and then we introduce a bagging model that bags the CNN model with two efficient SVM models reported in [4] to provide more robust prediction results. In particular, the contributions of this paper are as follows:

(1) We propose a novel fake review detection model, in which a CNN model is introduced to capture the product related review features and a classifier is established based on the product word composition features.

(2) To reduce overfitting and high variance of CNN model, we incorporate the CNN model with two

efficient SVM classification methods to build a bagging model for the classification task.

2. Related Work

Recently, many techniques and approaches have been proposed in the field of fake review detection. These methods exhibit high accuracy performance and can be roughly categorized as two categories: content based methods and behaviour feature based methods. We will illustrate these two kinds of methods in the following sections.

2.1. Content Based Method. Researchers attempt to distinguish review spam by analysing the contents of reviews, such as the linguistic features of the review [11]. To address the content feature of the reviews, Ott et al. checked three strateges to perform classification [4]. These three strategies are genre identification, detection of psycholinguistic deception, and text categorization [4, 11].

(i) *Genre Identification*. Ott et al. explored the parts-of-speech (POS) distribution of the review and use the frequency of POS tags as the features representing the review to make prediction.

(ii) *Detection of Psycholinguistic Deception*. The psycholinguistic technique is to assign psycholinguistic meanings to the key features of a review. Pennebaker et al. use the famous Linguistic Inquiry and Word Count (LIWC) software [12] to build their features for the reviews.

(iii) *Text Categorization*. According to the experiments of Ott et al., n-gram features play an important role at the experiments.

Other linguistic features are also explored, such as in the work [5]; Feng et al. take lexicalized and unlexicalized syntactic features using sentence parse trees for deception detection. Experiments show that the deep syntactic features improve the performance of prediction.

Li et al. [6] explored a variety of generic deceptive signals which contribute to the fake review detection. They also concluded that combine general features such as LIWC or POS with bag-of-words will be more robust than bag-of-words alone.

Metadata about reviews such as reviews length, date, time, and rating is also checked by some researchers [13, 14]. Experiments of their works show that the review characteristic features are beneficial in fake review detection.

Much of the previous work for fake review detection focused on related, but slightly different, issues, for example, using the linguistic features of review to detect fake reviews [4, 5] and exploring other features related to the reviews to build more efficient prediction models [6, 13, 14]. All these content based methods addressed detailed information closely related to the reviews. However, they paid little attention on the product related review features which is the main concerns of the proposed method.

2.2. Behaviour Feature Based Methods. Behaviour feature based models address the behaviour of individual reviewer, or groups of reviewers, including the "social relations" revealed by the reviewer behaviour.

Lim et al. identified the anomalous rating and review behaviours such as giving unfair ratings to products and reviewing too often, so as to detect spammers [7].

The works [7, 8] find that spammers may write fake reviews in collusion. Based on the findings, they make composed model to integrate these features for spammer detection.

Based on the network effect among reviewers and products, Akoglu et al. proposed a novel spammer and fake reviews spotting framework which is complementary to previous works based on text and behavioural features [9].

Fei et al. exploit the burstiness nature of reviews to spot review spammer [10]. Through a Markov Random Field model, their approach models the reviews in bursts and their cooccurrences in the same burst.

Since most of the above methods focus on analysing the behavioural features of the reviewers while the proposed method conducts the content of review, we will not compare the performance between our methods and theirs.

3. Validation of the Assumption of Product Related Review Feature

According to the observations of Li et al. [6], fake reviews have more positive/negative sentiment than the normal ones generated by actual customers. That is, review spammers emphasized some product features using more positive/negative words to agitate for/slander a product. This means that a particular product would be described by some special feature words and sentimental words when the spammers write the fake reviews. For example, product features in the hotel domain like the name of the hotels and the name of the staff and sentimental words like "extremely comfortable" are widely used [4]. In other domains, according to their findings [15], smartphone is often evaluated by "sleek" and "stable" and keyboard is evaluated by "wireless" and "mechanical." This product oriented information affects the performance of the prediction; thus integrating them into a classification model will benefit the classifier a lot.

To check the product related review features, we conduct the following experiments by using Algorithm 1 which is clearly discussed in the work [15]. To check the product related review features, we test it for $n = 100$ iterations on the dataset of Amazon product reviews [8]. In each iteration, reviews on the same product i (r_i, r_i^+) are first randomly sampled, and review r_i^- for other products is randomly chosen. After that, we calculate the similarity of (r_i, r_i^+) and (r_i, r_i^-), in which cosine similarity based on bag-of-words of two reviews is adopted.

As shown in Figure 1, the content similarities between two reviews about the same product are higher than those of different products (t-test with p value < 0.01). That is, the contents for the same product are more similar than for different products. This validates our assumption.

Input: review data R, number of products m, number of iterations n
Output: sim, dif
for $k = 1$ *to* n *do* **do**
  iSim = 0, iDif = 0;
  for $i = 1$ *to* m *do* **do**
    sample r_i, r_i^+, r_i^- from R;
    iSim += Similar(r_i, r_i^+);
    iDif += Similar(r_i, r_i^-);
  end
  iSim /= m, iDif /= m;
  sim ← sim ∪ iSim;
  dif ← sim ∪ iDif;
end
return sim, dif

Algorithm 1: Product related review features testing.

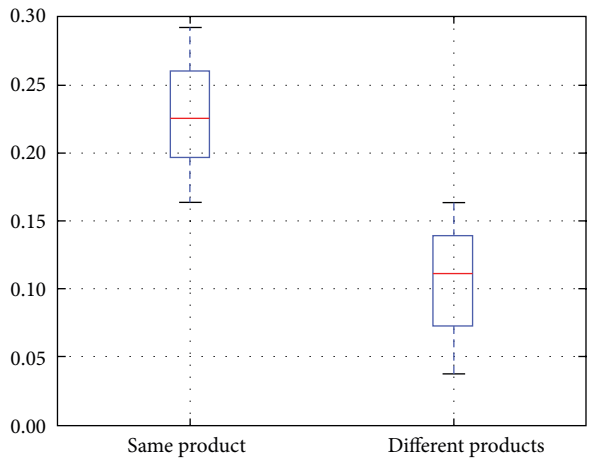

Figure 1: Validation of the assumption.

4. The Proposed Method for Fake Review Detection

In this section, we illustrate the proposed model for fake review detection in which we address the issue as a classification task. As shown in Figure 2, the proposed model accepts products and reviews as its input and generates classification results as its output. The proposed method offers classification results through a bagging model which bags three classifiers including product word composition classifier (PWCC), TRIGRAMS$_{SVM}$ classifier, and BIGRAMS$_{SVM}$ classifier. PWCC is a CNN model which captures product related review feature by a product word composition, so the product and review information can be fed into it for generating predictions. BIGRAMS$_{SVM}$ and TRIGRAMS$_{SVM}$ are two models reported in previous work to be efficient for prediction task. Both of them take the review as their input, and, in the proposed method, they are bagged with PWCC to produce more robust results.

In the following sections, we first illustrate PWCC in detail, and then we will introduce how to bag the three classifiers.

4.1. Product Word Composition Classifier. As discussed in Section 3, the deceptive reviews for every product have underlying relations with respect to the product. Thus we simply introduce a product word composition classifier to predict the polarity of the review. Following the ideas of [15], we first build a product-specific modification of the continuous representation of a word using the same way that Tang et al. model the user-specific modification. Then based on the output of the composition model, we build the document model and finally we use a CNN classifier to predict the reviews.

4.1.1. Product Word Composition. The product word composition model is used to map the words of a review into the continuous representation while concurrently integrating the product-review relations. In this paper, we employ the multiplicative composition to compose the product-specific modification. The multiplicative composition is detailed as follows. Given two vectors v_1 and v_2 as the input, multiplicative composition assumes that the output vector o is a linear function of tensor product of v_1 and v_2 which is shown as follows:

$$o = T \times v_1 \times v_2 = P_1 \times v_2. \quad (1)$$

Here, T is the tensor to project v_1 and v_2 to o. P_1 is the partial product of T and v_1. Based on (1), the multiplicative composition can exactly satisfy our requirements of modelling product-specific relations related to the reviews since the matrix P_1 models the products and v_2 illustrates the words in the reviews.

After conducting product word linear composition, we append tanh as the activation layer to integrate the nonlinearity attribute as shown in Figure 3. Hence, the final modified word vector o_i for the original word vector v_i is calculated as follows:

$$o_i = \tanh\left(w_{ik}\right) = \tanh\left(P_k \times v_i\right). \quad (2)$$

4.1.2. Document Modelling and Classification. To build the document model, we take the product word composition vectors as input and use CNN to build the representation model

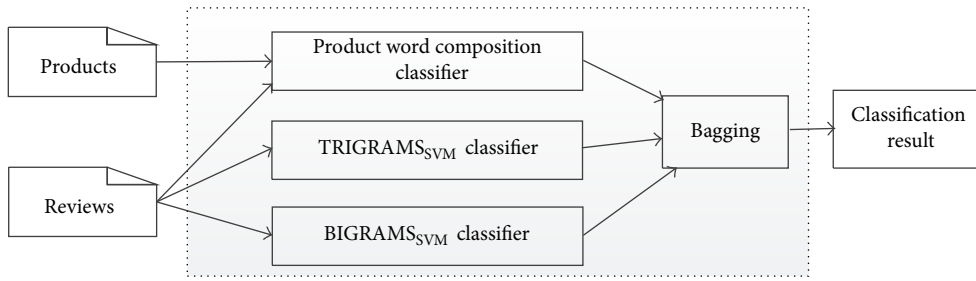

Figure 2: The proposed classification method.

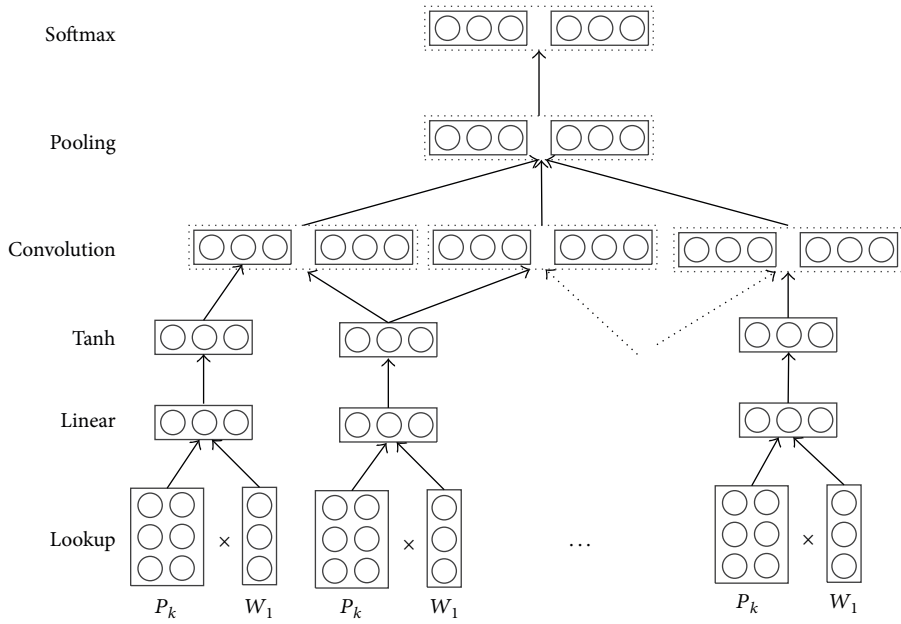

Figure 3: The product word composition classifier.

for the reviews. As shown in Figure 3, we feed product word composition vectors as the input of an average pooling layer to create the document model. Specifically, we use softmax to calculate the vector for the product word composition for generating the document vector as shown in the following:

$$\text{softmax}(\mathbf{x})_i = \frac{\exp(x_i)}{\sum_{j=1}^{C} \exp(x_j)}. \quad (3)$$

Here, C is the number of categories. Since the output of softmax can be interpreted as conditional probabilities, it is used to predict the polarity the reviews.

4.2. SVM Classifier and Bagging. As discussed above, we proposed a product word composition classifier to make prediction for deceptive reviews. However, the neural network model for this research may be overfitting and have high variance in the learned parameters over a little dataset. Specially in the research field of deceptive review detection, there are few good sources of labelled data [4]. Although more and more labelled data for this task has been published [6], it is not sufficient enough to fully take advantage of the power of deep learning model as the data is particular for classification for different domains. Therefore, it is helpful to build a model for alleviating this problem. In this paper, we use bagging method to deal with this issue, since the bagging method leads to "improvements for unstable procedures" [16], which is suitable for the neural networks. As discussed in Algorithm 2, we use bagging method to combine the product word composition based CNN model with two SVM models which have better precision for predicting the fake reviews according to this work [4].

Algorithm 2 bags these three classifiers to provide prediction results. It is composed of two phases: training and classification, respectively. In the first phase, three classifiers are trained using three bootstrap sample sets. Then, in the second phase, each input data is checked by all the classifiers in C, and the class label for each input with maximum number of votes is chosen.

Table 1: Statistics of the dataset.

Product	Number of reviews	Number of deceptive reviews	Number of truthful reviews
100	2000	800	1200

Training phase;
(1) Initialize the parameters
  (i) $\mathscr{C} = \emptyset$, the ensemble.
(2) **for** $k = 1, \ldots, 3$ **do**
  (i) Choose a bootstrap set S_k from Z.
  (ii) Build a classifier c_k using S_k.
  (iii) Add the classifier to the current ensemble, $\mathscr{C} = \mathscr{C} \cup \{c_k\}$
end
(3) **return** $\mathscr{C}$
Classification phase;
(4) Run $c_1, \ldots, c_L$ on the input x.
(5) The class with the maximum number of votes is chosen as the label for x.

Algorithm 2: Bagging the three classifiers.

5. Experiment

We conduct several experiments to evaluate the proposed model by applying it to reviews of products.

5.1. Experiment Setting. A gold-standard dataset [4] for fake review detection is widely used for validating different models. However, since it is argued that the fake reviews written by the Amazon Mechanical Turk are not reliable [17]. We attempted to create a dataset similar to the golden-standard dataset from the real-life dataset in [8] (http://liu.cs.uic.edu/download/data/). This dataset is about the reviews from amazon.com which is large and covers a very wide range of products. It is thus reasonable to consider it as a representative ecommerce site. The review dataset was crawled from amazon.com in June 2006. 5.8 million reviews, 2.14 reviewers, and 6.7 million products are included in this dataset. We created the dataset based on Amazon dataset using the following steps.

First, we use some seed words such as "full of fake reviews" to locate records of reviews. Depending on these reviews, we can find the products that the reviews relate to. This step is to find some products whose reviews may contain fake reviews since the reviews including seed words may be written by some users who are deceived to buy the product. Secondly, we remove the reviews with rating less than 4 and manually check whether the review is fake.

Using the above steps, we have collected 100 products where each product has 20 reviews. These 20 reviews are composed of 8 fake reviews and 12 truthful reviews. The statistic information of the dataset is shown in Table 1.

When training the CNN model, we split the data into training, validation, and testing sets with a 80/10/10 split and then split sentences and conduct tokenization with NLTK (http://www.nltk.org/). The two SVM based models are trained according to the configurations in [4].

When using the PWCC model, we set the widths of three convolutional filters as 1, 2, and 3. We learn 150-dimensional product-specific word embeddings on each dataset; other parameters are initialized randomly from a uniform distribution Uniform([0.01, 0.05]). The KISS random search for hyper parameters is adopted (http://deeplearning.net/tutorial/rnnslu.html#training).

To measure the overall classification performance, we use standard precision p, recall r, and f-measure f. Similarly, p, r, and f for the prediction are defined as follows:

$$p = \frac{|\text{golden} \cap \text{predicted}|}{|\text{predicted}|},$$
$$r = \frac{|\text{golden} \cap \text{predicted}|}{|\text{golden}|}, \quad (4)$$
$$f = 2 \cdot \frac{p \cdot r}{p + r},$$

where golden is the golden class labels and predicted is the predicted results of the classification methods.

5.2. Baseline Methods. We compare our method with the following baseline methods for review rating prediction:

(i) $\text{BIGRAMS}_{\text{SVM}}$: Ott et al. [4] propose to represent each review with bigrams feature set on which they train a SVM classifier for the fake review detection task.

(ii) $\text{TRIGRAMS}_{\text{SVM}}$: in this method, trigrams feature set is introduced to build the SVM classifier [4].

(iii) PWCC: we combine each review with the product to make a product word composition and then build a CNN classifier based on the composition for fake review prediction.

Table 2: Performance of the proposed model.

Methods	f	p	r
$\text{BIGRAMS}_{\text{SVM}}$	0.714	0.696	0.732
$\text{TRIGRAMS}_{\text{SVM}}$	0.722	0.703	0.741
PWCC	0.749	0.741	0.759
Bagging	0.772	0.764	0.781

(iv) Bagging: as discussed in Section 4, the bagging model combines the above three classifiers in order to offer more robust and accurate result.

5.3. Results and Analysis

5.3.1. Performance Analysis. Results appear in Table 2. After comparing the bagging method with the other models, we reach several important observations.

First, f, p, and r performance of the proposed bagging method outperforms the other methods from $\text{BIGRAMS}_{\text{SVM}}$ to PWCC. This demonstrates the effectiveness of the proposed method.

Second, there are little performance improvement from $\text{BIGRAMS}_{\text{SVM}}$ to $\text{TRIGRAMS}_{\text{SVM}}$. This reveals that the contributions of linguistic features will be limited after reaching an upper bound. Combining with other features may alleviate the problem and contributes to getting better performance.

Third, the performance of PWCC performs better than both $\text{BIGRAMS}_{\text{SVM}}$ and $\text{TRIGRAMS}_{\text{SVM}}$. This improvement of performance of PWCC may be due to two reasons: one is that the CNN model has better prediction performance than the SVM based model. The other reason may be that composition of product and word contributes to the better results.

5.3.2. Analysis of Product Word Composition. We investigate the effects of product word composition model which integrates product related review features for fake review detection. Since the product word composition is composed of product and word information, we remove the representations P from PWCC model to build a CNN classifier based on word representation and then conduct experiments on Amazon dataset.

As shown in Figure 4, we can see that PWCC achieves better results of f, p, and r. Compared with PWCC, the CNN model only using word features removed the product related composition information. This means the improvement of performance is mainly brought by adding composition information.

6. Analysis of Classifiers

To find which algorithm outperforms others on the learning task in this paper, we introduced $5*2$ cv test which is based on 5 iterations of 2-fold cross-validation according to Dietterich's work [18].

Figure 5 shows the measured Type 1 error rates of the four methods used in this paper. As shown in Figure 5, we can see that *bagging* achieves better results of lower probability of Type 1 error. This means bagging all the three methods brings improvement of robustness for avoiding Type 1 error.

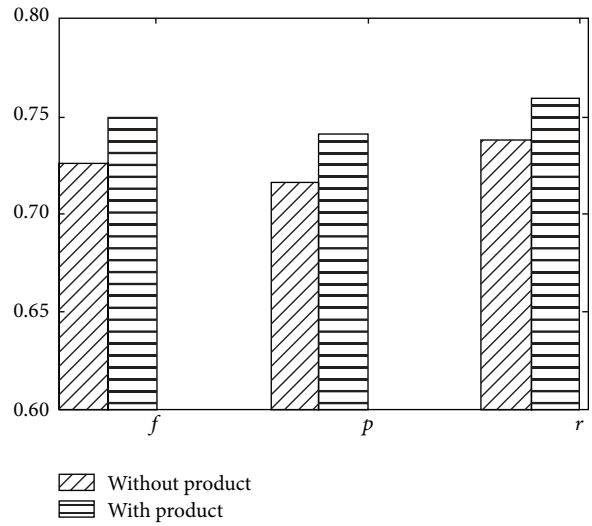

Figure 4: Analysis of product word composition.

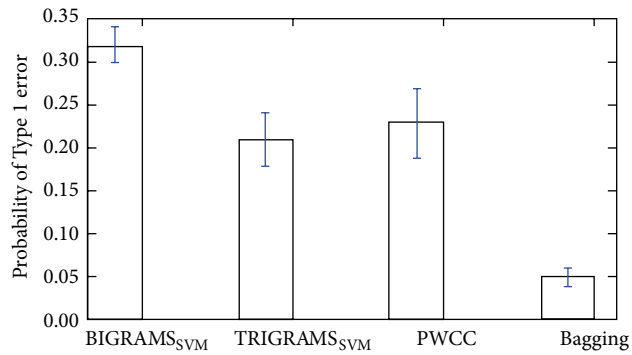

Figure 5: Analysis of test.

7. Conclusion

This paper exploits the product related review features for fake review detection. A novel convolutional neural network model is proposed to composite the product and word feature. To provide reduced overfitting and high variance, we use bagging strategy to bag the neural network model with two efficient classifiers. To evaluate the proposed method, we attempted to create a dataset from a real-life review dataset. A variety of experiments are conducted to analyse the effectiveness of the proposed model.

However, there exist other kinds of review or reviewer related features that are likely to make a contribution to the prediction task. In the future, we could further investigate different kinds of features to make more accurate predictions.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The work is supported by the National Basic Research Program of China under Grant no. 2014CB340404, University of Science and Technology Program of Shandong Province under Grant no. J16LN08, Scientific Research Foundation of Shandong University of Science and Technology for Recruited Talents under Grant no. 2016RCJJ045, the State Key Laboratory of Software Engineering Foundation under Grant no. SKLSE 2014-10-07, University Teaching Reform Project of Shandong Province under Grant no. 2015M140, and Educational Science Research of Shandong Province under Grant no. 15SC111.

References

[1] Ipsos, "Socialogue: five stars? thumbs up? a+ or just average?" 2012.

[2] M. Ott, C. Cardie, and J. Hancock, "Estimating the prevalence of deception in online review communities," in *Proceedings of the 21st Annual Conference on World Wide Web (WWW '12)*, pp. 201–210, Lyon, France, April 2012.

[3] A. Mukherjee, "Detecting deceptive opinion spam using linguistics, behavioral and statistical modeling," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, Technical Report, Beijing, China, July 2015.

[4] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT '11)*, vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011.

[5] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers—Volume 2*, pp. 171–175, Association for Computational Linguistics, 2012.

[6] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL '14)*, pp. 1566–1576, June 2014.

[7] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10)*, pp. 939–948, ACM, Toronto, Canada, 2010.

[8] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the International Conference on Web Search and Data Mining (WSDM '08)*, pp. 219–230, ACM, February 2008.

[9] L. Akoglu, R. Chandy, and C. Faloutsos, "Opinion fraud detection in online reviews by network effects," *ICWSM*, vol. 13, pp. 2–11, 2013.

[10] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting burstiness in reviews for review spammer detection," *ICWSM*, vol. 13, pp. 175–184, 2013.

[11] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, "Detection of review spam: a survey," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3634–3642, 2015.

[12] J. W. Pennebaker, M. E. Francis, and R. J. Booth, *Linguistic Inquiry and Word Count: Liwc 2001*, vol. 71, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2001.

[13] F. Li, M. Huang, Y. Yang, and X. Zhu, "Learning to identify review spam," in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI '11)*, pp. 2488–2493, Barcelona, Spain, July 2011.

[14] A. A. Hammad and A. El-Halees, "An approach for detecting spam in arabic opinion reviews," *International Arab Journal of Information Technology*, vol. 12, no. 1, pp. 10–16, 2015.

[15] D. Tang, B. Qin, and T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Volume 1: Long Papers (ACL-IJCNLP '15)*, pp. 1014–1023, Beijing, China, July 2015.

[16] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[17] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "Fake review detection: classification and analysis of real and pseudo reviews," Tech. Rep. UIC-CS-2013-03, University of Illinois at Chicago, Chicago, Ill, USA, 2013.

[18] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Computation*, vol. 10, no. 7, pp. 1895–1923, 1998.