

Learning Document Representation for Deceptive Opinion Spam Detection

Luyang Li, Wenjing Ren, Bing Qin*, and Ting Liu

Research Center for Social Computing and Information Retrieval,
Harbin Institute of Technology, Harbin, China
{lyli, wjren, qinb, tliu}@ir.hit.edu.cn

Abstract. Deceptive opinion spam in reviews of products or service is very harmful for customers in decision making. Existing approaches to detect deceptive spam are concern on feature designing. Hand-crafted features can show some linguistic phenomenon, but is time-consuming and can not reveal the connotative semantic meaning of the review. We present a neural network to learn document-level representation. In our model, we not only learn to represent each sentence but also represent the whole document of the review. We apply traditional convolutional neural network to represent the semantic meaning of sentences. We present two variant convolutional neural-network models to learn the document representation. The model taking sentence importance into consideration shows the better performance in deceptive spam detection which enhances the value of F1 by 5%.

Keywords: Deceptive Spam Detection, Convolutional Neural Network, Representation Learning

1 Introduction

Deceptive opinion spam detection is an urgent and meaningful task in the field of natural language processing. By continuous growth of the user-generated reviews, the appearance of deceptive opinion spam arouses people's attention [25, 24, 40, 42]. Deceptive opinion spam is the review with fictitious opinions which is deliberately written to sound authentic [34]. For commercial motive, some businesses hire people to write undeserving positive reviews to promote the products or giving unjust negative reviews to damage the reputations of the objects [14]. It is very difficult to distinguish deceptive spam by people. In the test of Ott et al. [34], the average accuracy of three human judges is only 57.33%. Hence, the research in detecting deceptive opinion spam automatically by machine is necessary.

The review is always a short document consisting of a few sentences. The objective of the task is to distinguish the document whether a spam or a truth. The task can be transformed into a spam classification problem. The majority of existing approaches follow Ott et al. [34] and employ machine learning algorithms to build classifiers. Under this direction, most studies focus on designing effective features to better classification performance. Feature engineering is important but labor-intensive. It also can

* Corresponding Author

not reveal inherent law from the semantic perspective in data. For the task of deceptive opinion spam detection, an effective feature learning method is to compose the representation of the document. Learning the representation of the document can capture the global feature and take word order and sentence order into consideration. That has more advantages than common features like n-grams, POS, etc.

We aim to learn the representation of the document for deceptive opinion spam detection. The learning model is consisting of two stages which are sentence representation learning and document representation learning. At the stage of sentence representing, we apply sliding window to capture sequential words and transform to a vector. We exploit two variant models of convolution neural network to learn document representation which are different at the second stage. In consideration of the effect of sentence order to semantic representing, our first model, sentence convolutional neural network (SCNN), apply sliding window to capture sequential sentences and transform to a vector. Namely, a multilayer convolutional neural network is applied to learn the representation of the document. In a review, which is a document, a few sentences may include the more important concepts, and thus should be more heavily weighted. Based on the consideration, we utilize information gain to evaluate the importance of sentences, and develop a sentence weighted neural network (SWNN) by assigning a different weight according to the importance of the sentence to each term.

We use a basic method to represent document and apply as features in a supervised learning framework for deceptive opinion spam detection on the public data sets [21] and gain an comparable result with state-of-the art method. We also find that our document representation perform more robust on cross-domain data. We also apply the two variant models of convolution neural network on the mixture-domain data sets, and SWNN model gains better performance than baseline methods. The major contributions of the work presented in this paper are as follows.

- We show that the document representation based on the word embedding performs more robust than traditional common feature on cross-domain data in the task.
- We exploit two convolutional neural-network based models to learn the document representation and the results show the effectiveness on the public data sets.

2 Methodology

In the section, we present the details of learning document representation for deceptive opinion spam detection. We extend the existing text representation learning algorithm [4] and develop two convolutional neural network models to learn document representation for deceptive opinion spam detection. In the following subsections, we first introduce the traditional method and then present the detail of two document representation learning models.

2.1 Basic Convolutional Neural Network

Collobert et al. [4] introduce a sentence approach network to learn the representation of a sentence. The architecture is given in Figure 1. It is a multilayer neural network

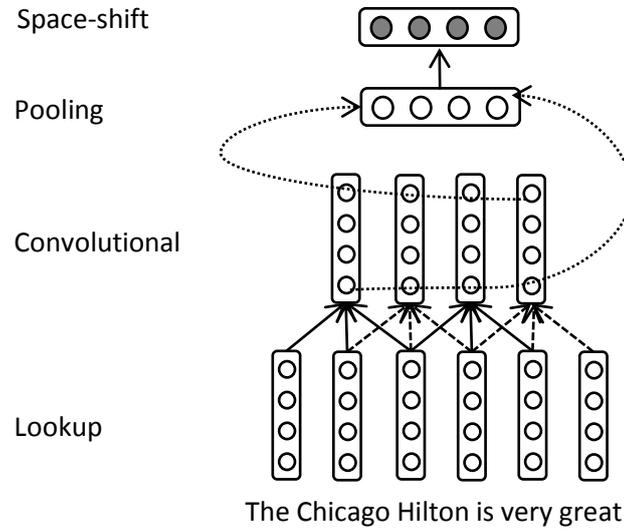


Fig. 1. The traditional neural network for learning sentence representation.

which consists of four types of layers. Given a sentence “*The Chicago Hilton is very great*”, the model apply the lookup layer to map these words into corresponding word embeddings which are continuous real-valued vectors. The convolutional layer extracts local features around each window of the given sequence by representing the semantic meaning of the words in the window. The size of the output of convolutional layer depends on the number of words in the sentence fed to the network. Pooling layer obtain a global feature vector by combining the local feature vectors through previous layer. Common operations are doing max or average over the corresponding position of the sequence. The average operation captures the influence of all words to the certain task. The max operation captures the most useful local features produced by convolutional layer. The space-shift layer include linear layer and non-linear layer, and maybe include another linear layer if the output is scores of corresponding categories in certain task. Non-linear layer is necessary to extract high level features.

2.2 The document representation learning model

Basic model. We apply the traditional convolutional neural network model to represent sentences. To make a composition for the document, we use average operation to capture all of the sentences features on the pooling layer. This is a basic model, which is modified below to suit the deceptive opinion spam detection task.

SCNN model. As the architecture is given in the Figure 2, SCNN model consists of two convolutional layers to do the composition. The sentence convolution is to make a composition of each sentence by a fix-length window. The document convolution transforms sentence vectors into a document vector. The ranking layer produces the

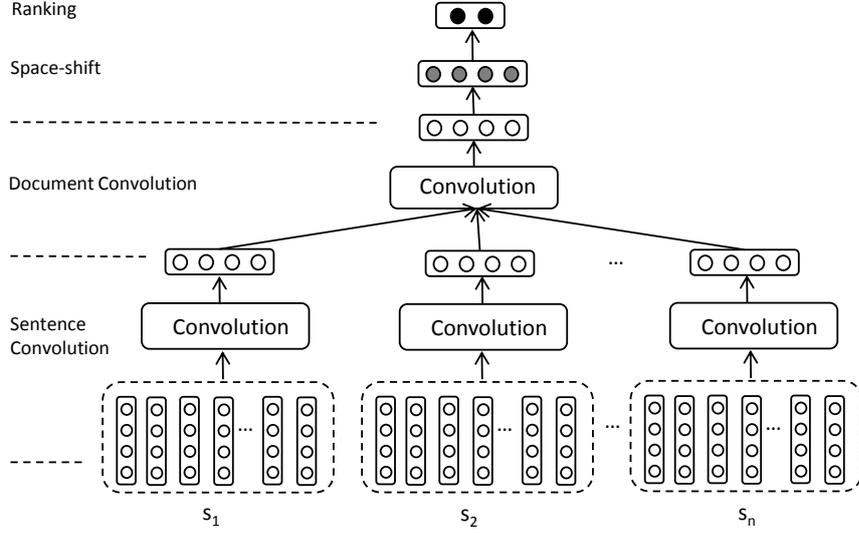


Fig. 2. The our SCNN model for learning sentence representation.

scores according to each category. We use hinge loss as the ranking objective function in Equation 1.

$$Loss(r) = \max(0, m_\delta - f(r_t) + f(r_{t^*})) \quad (1)$$

Where t is the gold label of the review r , t^* stands for the another label, m_δ is the margin in the experiment.

SWNN model. The sentence-weighted neural network model is a modified model of the basic document representation learning model. As a matter of fact, the words in a review play different roles in the semantic representation. Some words must be more important in distinguishing spam from the truth reviews. Hence, each sentence also owns its importance weight according to the words in it. We compute the importance weight of the sentence based on the importance weights of words in the sentence. We apply KL-divergence as the importance weight of the word. The value of KL-divergence stands for the capacity of a feature in dividing documents which is a feature selection approach. In fact, we also try $tf - idf$ as a candidate of weight computing method, however, it does not perform as well as KL-divergence in the experiment. We assume that $U = \{U_1, \dots, U_i, \dots, U_n\}$ is the universal set of words in the review which U_i is the word set of the i th sentence, and W_j stand for the weight of the j th word. The sentence weight is a normalization value like in the following formula.

$$\alpha_i = \frac{\sum_{j \in U_i} W_j}{\sum_{k \in U} W_k} \quad (2)$$

In the Figure 3, the architecture of SWNN model is given. Each sentence of the input document review transforms into the fixed-length vector through convolutional

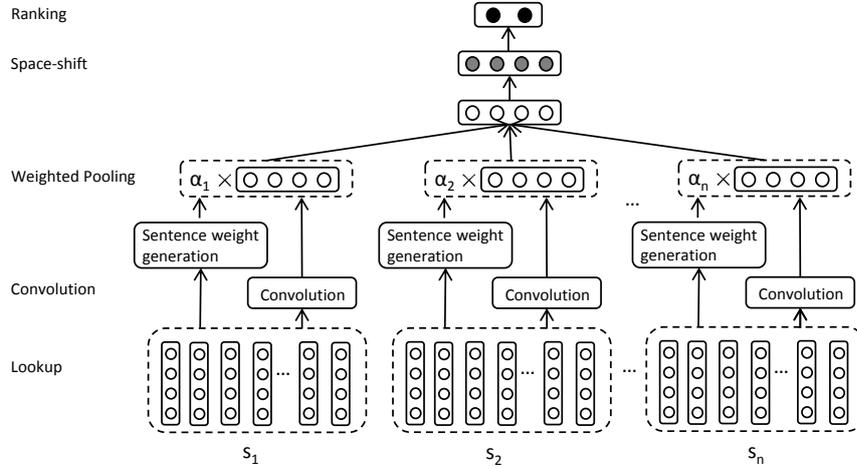


Fig. 3. The our SWNN model for learning sentence representation.

layer. The process of generating sentence weights produce normalized weight α_i corresponding to the i th sentence. Through the pooling layer, the sentence vectors transform into a document vector by a weighted-average operation. More important sentences have more influences during producing the document vector. The vector transforms through space-shift layer to extract high level features. The ranking layer produce the scores of the categories.

3 Experiment

We conduct experiments to empirically evaluate our document representation learning model by applying the it to do the deceptive opinion spam detection task. We do two comparison experiments to show the effectiveness our model.

3.1 Experiment Setup

We apply the public data sets released by Jiwei Li [21]. The data sets contain three domains which are hotel, restaurant, and doctor. The distribution of the dataset is shown in Table 1. In Li’s public dataset, there are three types of data in each domain which are “Turker”, “Expert”, and “Customer”. They stand for various different data sources. The spam reviews are edited by Turkers and experts who have domain knowledge. The truth reviews are from customers who really have consumption experience. However, Li do not apply “Expert” data in his experiment. According to Li’s paper, he only apply 200 spam reviews from 356 spam reviews in Doctor domain. Hence, we do our best to use data with the same distribution in the cross-domain experiment comparing with Li.

Our target is to exploit domain-independent method to resolve deceptive opinion spam detection. Hence, we construct a mixture domain dataset. The samples in the

	Turker	Customer
Hotel(P/N)	400/400	400/400
Restaurant(P/N)	200/0	200/200
Doctor(P/N)	356/0	200/0

Table 1. Statistics of the three domain dataset.

dataset are divided into two categories, i.e. spam (Turker) and truth (Customer). The proportion among training set, development set and test set is 6 : 1 : 3. Each category data in each domain is assigned by the proportion.

3.2 Cross-domain Classification

To frame the problem as a domain adaptation task, we want to find a more robust feature on cross-domain dataset. On the latest public data, only Li show the experiment results. Hence, we do the comparison with his method. We apply basic document representation as features which is the average vector of all word embedding in the paragraph.

Baseline Method. Li respectively apply Unigram, LIWC and POS features in SVM and SAGE classifiers to explore a more general classifier of the task. SAGE is sparse additive generative model which can be viewed as an combination of topic models and generalized additive models. However SAGE do not outperform SVM, we apply SVM as the classifier in the comparison experiment. In Li’s experiment, the method gains best results by using Unigram an POS features in test datasets (restaurant and doctor domains) by training hotel domain data. Hence, we just list the best results from his paper.

Features	Restaurant				Doctor			
	A	P	R	F1	A	P	R	F1
Unigram	0.785	0.813	0.742	0.778	0.550	0.573	0.725	0.617
POS	0.735	0.697	0.815	0.751	0.540	0.521	0.975	0.679
Paragraph-average	0.733	0.684	0.865	0.764	0.588	0.555	0.885	0.682

Table 2. Classifier performance in cross-domain test data.

Results and Analysis. Table 2 show the results from baseline method as well as our method. We can see the our basic document representation perform comparable respectively with the best results of baseline on two domain. Additionally, the document representation perform more robust on the cross-domain dataset.

3.3 Domain-independent Classification

We apply various document representations learnt by our variant neural network models as features to do the deceptive spam classification. As we introduce above, we randomly

construct domain-independent datasets by the uniform distribution from three domain data. For each variant model, we train on the training set, adjust parameters on the development set and predict on the test set.

The Basic CNN is the basic convolutional neural network model which sentences are representing through convolutional layer and transform into a document vector by the average operation. SCNN apply convolutional layer to replace the average operation . SWNN is the modification of the Basic CNN model by using sentence weights.

Model	A	P	R	F1
Paragraph-average	0.676	0.673	0.903	0.771
Basic CNN	0.709	0.739	0.813	0.774
SCNN	0.699	0.670	0.859	0.753
SWNN	0.795	0.761	0.898	0.823

Table 3. Deceptive opinion spam classification.

Results and Analysis. We do the comparison among various document representations. Table 3 show the results that our SWNN model learn the best representation and gain the best result in deceptive spam classification. The scores of accuracy and F1 are all far above the other neural-network based methods. The results show the effectiveness of incorporating sentence weight in representing document. We also find more complex model like SCNN do not perform as well as simple model like Paragraph-average model and Basic CNN model.

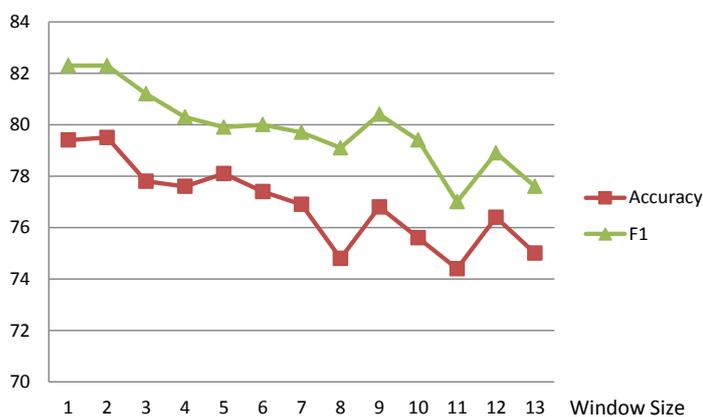


Fig. 4. Effect of window size.

Parameter Settings. The parameters of SWNN model used in the deceptive opinion spam detection experiment is listing followed. The embedding length and the vector length in two hidden layers are all 50. The learning rate is 0.1. The window size is set as 2. We experimentally study the effect of window size in our presented convolutional neural network method. We tune the parameter on trial dataset. In Figure 4, we vary the value of window size and compute the accuracy and F1. It shows the accuracy scores have one top (0.795) when the value of window size equals to 2 which we applies in the test. The F1 also has the best result at the same point.

4 Related Work

We present a brief review of the related work from two perspectives. One is deceptive opinion spam detection, and another is deep learning for specific task representation learning.

4.1 Deceptive Opinion Spam Detection

On the Internet, various kinds of spam brings troubles to people. Over the years, many studies focus on spam detection. Web spam has been extensively studied [11, 30, 10, 23, 45, 8, 2] The objective of the web spam is to gain high page rank and attract people to click by fooling search engines. Email spam is another related research, which is pushing unsolicited advertisements to users [3, 5]. The web spam and mail spam have a common character that they have irrelevant words. Opinion spam is quite different and more crafty. By the explosive growth of user-generated content, the number of opinion spam in the reviews, which contain opinions of users about products and services, increased continuously. This phenomenon attracted researchers attention. Opinion spam was firstly investigated by Liu et al. [14] that also summarized the opinion spam into different types. In terms of the different damage to users, we can further conclude the opinion spam into two types which are deceptive opinion spam and product-irrelevant spam. In the former spam, the spammers give undeserving positive reviews or unjust negative reviews to the object for misleading costumers. The latter spam contain no comments about the object. Obviously, the deceptive opinion spam is more difficult to detect.

The approaches of detecting deceptive opinion spam can be divided into unsupervised methods and supervised methods. Liu et al. [27] take a Bayesian approach and formulate opinion spam detection as a clustering problem. There are also many unsupervised methods researching on detecting spammers [44, 29, 22, 28] or mining reviewing patterns [15]. Due to the lack of gold standard data, most methods take the research on pseudo labeled data. Liu et al. [14] assumed duplicate and near duplicate reviews to be deceptive spam. They also applied features of review texts, reviewers and products. Yoo et al. [47] first collected a small amount of deceptive spam and truth reviews and do a linguistic analysis on them. By applying Amazon Mechanical Turk, Ott et al. [34, 32, 33, 31] gathered a gold standard labeled data. A few follow-up researches have been done on the data set. Ott et al. estimated prevalence of deceptive opinion

spam in reviews [32], and identified negative spam [33]. Li et al. [20] identified manipulated offerings on review portals. Feng et al. [6] applied context free grammar parse trees to extract syntactic features to improve the performance of the model. Vanessa Feng et al. [7] take the group of reference reviews into account according to the same product. Although there are deceptive opinion spam in the Ott’s data sets, it still can not reflect the real condition with the lack of cross-domain data, and the Turkers also lack of professional knowledge. Li et al. [21] created a cross-domain data sets (i.e. hotel, restaurant, and doctor) with part of reviews from domain experts. On this labeled data set, they use n-gram features as well as POS and LIWC features in classification and show that POS perform more robust on cross-domain data.

4.2 Deep Learning for Representation Learning

Representation learning by deep learning methods has been proven to be effective in avoiding task-specific engineering. Hence, the processing does not need much prior knowledge. As a continuous real-valued vector, representation can be incorporated as features in a variety of natural language processing tasks [4, 16, 19], such as POS tagging, chunking, named entity recognition [4, 43], semantic role labeling, parsing [36], language modeling [26, 1], and sentiment analysis tasks [39, 41]. Representation learning is to learn continuous representations of text with different grains, like word, phrase, sentence and document. For representing a document, the existing deep learning methods consist of two processing stages. Firstly, word embedding should be learnt by massive text corpus. Some work utilizes global context of document and multiple word prototypes [13], or global word-word co-occurrence to improve word embedding [35]. There are also some work for task-specific word-embedding [41]. After obtaining word representation, many research works focus on composing for coarse-grained semantic unit by composition models. For learning semantic composition,

Yessenalina et al. use matrixes to model each word and applying iterated matrix multiplication to combine words [46]. Glorot et al. develop Stacked Denoising Autoencoders for domain adaptation [9]. Socher et al. propose Recursive Neural Network (RNN) [38], matrixvector RNN [37] and Recursive Neural Tensor Network (RNTN) [39] to learn the compositionality of unfixed-length phrases. Hermann et al. (2013) learn the compositionality of sentence by Combinatory Categorical Autoencoders, which is the combination of Combinatory Categorical Grammar and Recursive Autoencoder [12]. Li et al. [18] use feature weight tuning to control the effect one specific unit makes to the higher-level representation in a Recursive Neural Network. Le et al. [17] learn the representation of paragraph.

5 Conclusion

We introduce a novel convolutional neural network to learn document representation for deceptive opinion spam detection. Sentences play different important role in the document. We model the semantic meaning of document-level reviews by incorporating sentence important weights into document representation learning. We construct experiments on the latest public data set and compare with multiple baseline methods.

We show that sentence-weighted neural network is more effective than other two convolutional neural-network based models in document representation and spam classification. The results of the experiments also show that the basic document representation perform more robust than the hand-crafted features on cross-domain data set.

Acknowledgments. This work was supported by the National High Technology Development 863 Program of China (NSFC) via grant 2015AA015407, NSFC via grant 61133012 and NSFC via grant 61273321.

References

1. Bengio, Y., Ducharme, R., Vincent, P., Janvin, C.: A neural probabilistic language model. *The Journal of Machine Learning Research* 3, 1137–1155 (2003)
2. Castillo, C., Donato, D., Gionis, A., Murdock, V., Silvestri, F.: Know your neighbors: Web spam detection using the web topology. In: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. pp. 423–430. ACM (2007)
3. Chirita, P.A., Diederich, J., Nejdl, W.: Mailrank: using ranking for spam detection. In: *Proceedings of the 14th ACM international conference on Information and knowledge management*. pp. 373–380. ACM (2005)
4. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. *The Journal of Machine Learning Research* 12, 2493–2537 (2011)
5. Drucker, H., Wu, D., Vapnik, V.N.: Support vector machines for spam categorization. *Neural Networks, IEEE Transactions on* 10(5), 1048–1054 (1999)
6. Feng, S., Banerjee, R., Choi, Y.: 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)
7. Feng, V.W., Hirst, G.: Detecting deceptive opinions with profile compatibility. In: *Proceedings of the 6th International Joint Conference on Natural Language Processing, Nagoya, Japan*. pp. 14–18 (2013)
8. Fetterly, D., Manasse, M., Najork, M.: Detecting phrase-level duplication on the world wide web. In: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. pp. 170–177. ACM (2005)
9. Glorot, X., Bordes, A., Bengio, Y.: Domain adaptation for large-scale sentiment classification: A deep learning approach. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. pp. 513–520 (2011)
10. Gyöngyi, Z., Garcia-Molina, H.: Link spam alliances. In: *Proceedings of the 31st international conference on Very large data bases*. pp. 517–528. VLDB Endowment (2005)
11. Gyöngyi, Z., Garcia-Molina, H., Pedersen, J.: Combating web spam with trustrank. In: *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*. pp. 576–587. VLDB Endowment (2004)
12. Hermann, K.M., Blunsom, P.: The role of syntax in vector space models of compositional semantics. In: *ACL (1)*. pp. 894–904 (2013)
13. Huang, E.H., Socher, R., Manning, C.D., Ng, A.Y.: Improving word representations via global context and multiple word prototypes. In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. pp. 873–882. Association for Computational Linguistics (2012)

14. Jindal, N., Liu, B.: Opinion spam and analysis. In: Proceedings of the 2008 International Conference on Web Search and Data Mining. pp. 219–230. ACM (2008)
15. Jindal, N., Liu, B., Lim, E.P.: Finding unusual review patterns using unexpected rules. In: Proceedings of the 19th ACM international conference on Information and knowledge management. pp. 1549–1552. ACM (2010)
16. Kalchbrenner, N., Grefenstette, E., Blunsom, P.: A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188 (2014)
17. Le, Q.V., Mikolov, T.: Distributed representations of sentences and documents. arXiv preprint arXiv:1405.4053 (2014)
18. Li, J.: Feature weight tuning for recursive neural networks. arXiv preprint arXiv:1412.3714 (2014)
19. Li, J., Jurafsky, D., Hovy, E.: When are tree structures necessary for deep learning of representations? arXiv preprint arXiv:1503.00185 (2015)
20. Li, J., Ott, M., Cardie, C.: Identifying manipulated offerings on review portals. In: EMNLP. pp. 1933–1942 (2013)
21. Li, J., Ott, M., Cardie, C., Hovy, E.: Towards a general rule for identifying deceptive opinion spam
22. Lim, E.P., Nguyen, V.A., Jindal, N., Liu, B., Lauw, H.W.: Detecting product review spammers using rating behaviors. In: Proceedings of the 19th ACM international conference on Information and knowledge management. pp. 939–948. ACM (2010)
23. Metaxas, P.T., DeStefano, J.: Web spam, propaganda and trust. In: AIRWeb. pp. 70–78 (2005)
24. Meyer, D.: Fake reviews prompt belkin apology. CNet News (2009)
25. Miller, C.: Company settles case of reviews it faked. New York Times (2009)
26. Mnih, A., Hinton, G.E.: A scalable hierarchical distributed language model. In: Advances in neural information processing systems. pp. 1081–1088 (2009)
27. Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M., Ghosh, R.: Spotting opinion spammers using behavioral footprints. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 632–640. ACM (2013)
28. Mukherjee, A., Liu, B., Glance, N.: Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st international conference on World Wide Web. pp. 191–200. ACM (2012)
29. Mukherjee, A., Liu, B., Wang, J., Glance, N., Jindal, N.: Detecting group review spam. In: Proceedings of the 20th international conference companion on World wide web. pp. 93–94. ACM (2011)
30. Ntoulas, A., Najork, M., Manasse, M., Fetterly, D.: Detecting spam web pages through content analysis. In: Proceedings of the 15th international conference on World Wide Web. pp. 83–92. ACM (2006)
31. Ott, M.: Computational linguistic models of deceptive opinion spam (2013)
32. Ott, M., Cardie, C., Hancock, J.: Estimating the prevalence of deception in online review communities. In: Proceedings of the 21st international conference on World Wide Web. pp. 201–210. ACM (2012)
33. Ott, M., Cardie, C., Hancock, J.T.: Negative deceptive opinion spam. In: HLT-NAACL. pp. 497–501 (2013)
34. Ott, M., Choi, Y., Cardie, C., Hancock, J.T.: 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-Volume 1. pp. 309–319. Association for Computational Linguistics (2011)
35. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. Proceedings of the Empirical Methods in Natural Language Processing (EMNLP 2014) 12, 1532–1543 (2014)

36. Socher, R., Bauer, J., Manning, C.D., Ng, A.Y.: Parsing with compositional vector grammars. In: *Proceedings of the ACL conference*. Citeseer (2013)
37. Socher, R., Huval, B., Manning, C.D., Ng, A.Y.: Semantic compositionality through recursive matrix-vector spaces. In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. pp. 1201–1211. Association for Computational Linguistics (2012)
38. Socher, R., Lin, C.C., Manning, C., Ng, A.Y.: Parsing natural scenes and natural language with recursive neural networks. In: *Proceedings of the 28th international conference on machine learning (ICML-11)*. pp. 129–136 (2011)
39. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*. vol. 1631, p. 1642. Citeseer (2013)
40. Streitfeld, D.: For 2 a star, an online retailer gets 5 star product reviews. *New York Times* 26 (2012)
41. Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., Qin, B.: Learning sentiment-specific word embedding for twitter sentiment classification. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. vol. 1, pp. 1555–1565 (2014)
42. Topping, A.: Historian orlando figes agrees to pay damages for fake reviews. *The Guardian* 16 (2010)
43. Turian, J., Ratinov, L., Bengio, Y.: Word representations: a simple and general method for semi-supervised learning. In: *Proceedings of the 48th annual meeting of the association for computational linguistics*. pp. 384–394. Association for Computational Linguistics (2010)
44. Wang, G., Xie, S., Liu, B., Yu, P.S.: Review graph based online store review spammer detection. In: *Data mining (icdm), 2011 IEEE 11th international conference on*. pp. 1242–1247. IEEE (2011)
45. Wu, B., Davison, B.D.: Identifying link farm spam pages. In: *Special interest tracks and posters of the 14th international conference on World Wide Web*. pp. 820–829. ACM (2005)
46. Yessenalina, A., Cardie, C.: Compositional matrix-space models for sentiment analysis. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. pp. 172–182. Association for Computational Linguistics (2011)
47. Yoo, K.H., Gretzel, U.: Comparison of deceptive and truthful travel reviews. *Information and communication technologies in tourism 2009* pp. 37–47 (2009)