ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

A Comparative Study on Fake Review Detection Techniques

[1]A.Lakshmi Holla, [2] Dr Kavitha K.S
[1] Research Scholar, Computer Science and Engineering, Global Academy Of Technology, Bangalore
[2] Professor, Department of Computer Science and Engineering Global Academy of Technology, Bangalore

Abstract: **- Opinion mining has played a significant role in providing product recommendations to users. Efficient recommendation systems help in improving business and also enhance customer satisfaction. The credibility of purchasing a product highly depends on the online reviews. However many people wrongly promote or demote a product by buying and selling fake reviews. Many websites have become source of such opinion spam. This in turns leads to recommending undeserving products. This literature survey is done to study the various fake review detection techniques in detail and to get ourselves familiar with the works done on this subject.**

Index Term**- Fake Review Detection Systems, Opinion Mining, Recommendation Systems.**

I. INTRODUCTION

With the advent of huge amount of information available in the cyber space and large number of users , using this information, extracting the opinions of users has become crucial in the current world. Analyzing the user reviews and providing suitable recommendations to users based on this reviews has become significant in enabling the users to get the right information about a deserving product which they want to purchase. However it is often found that the efficiency of the recommendations systems is hampered due to its reliance on spam reviews posted by impostors whose main intention might be to demote or promote a product for profit. Spam reviews has become one of the major concerns and a big threat in todays world. There is a need for designing efficient fake review detection systems to tackle the menace of abundant fake reviews available on the web.

A. Characteristics of Fake Reviews

Some of the typical characteristics of fake reviews are:
1.Less Information about the Reviewer: Users who have less social connections and who do not have profile information are usually impostors and are likely to post only a few reviews which are fake reviews.
2. Review Content Similarity: Spammers often copy their own reviews or reviews of other users and often write duplicate or near duplicate reviews. These reviews may indicate spam reviews.
3. Short Reviews :As spammers are interested in making quick profits they tend to write very short reviews with a lot of grammatical errors and excessively use capitals, numerals and all capital words .They also focus on brand names of a product.
4. Sudden uploading of reviews in the same time frame: It is found that one of the best ways to detect fake reviews is by looking at the timestamp of the reviews and if a batch of reviews is uploaded at the same time then this is a spam indicator.
5. Focus on personal information: Genuine reviews usually focus on spatial information but spam reviews focus on irrelevant personal information such as people.
6. Excessive use of positive and negative words: Spammers often use a lot of positive and negative words in a review which might not be the necessity in the context of the review. Further they also write reviews based on the product description and not on their experience of using the product.

II. RELATED WORK

A. Detection of Fake Review created by Groups

Group spamming refers to a group of reviewers who work in collaboration and write reviews to promote or demote the reputation of a product. In[1] the author uses a frequent itemset mining method to find a set of candidate groups. Here frequent itemset refers to the set of multiple products reviewed by the users together. Further a minimum support count of three is set which indicates each group should have worked together on atleast three products together. Next the degree of spamicity was calculated for each group by assigning 1

point for spam reviews ,0 point for non-spam reviews and 0.5 for each borderline judgement. This was used as the basis for ranking the spam groups. Several group and individual behavioural spam indicators such as Group Time Window, Group size, Group Early Time Frame ,Group content similarity,Individual Time Window, Individual Content Similarity, Individual Early Time Frame are used to derive three relational models namely Group-Spam Products, Individual-Spam products and Group Spam-Member Spam. Finally an iterative ranking algorithm called GSRank (Group Spam Ranking Algorithm) is executed to draw

inferences from the three relational models and classify spam reviews. This algorithm outperforms all baseline methods like regression, rank and classification algorithms. However here the author has focused only on the behavioral parameters in detecting spam reviews and has not considered the effect of review text characteristics such as implicit sentiments of reviews, effect of modifiers and contextual use of sentiment words in fake reviews.

B. Generation of synthetic reviews and their detection

In this paper[2] the author discusses the generation of synthetic reviews by considering a truthful review as a template and replacing the sentences of this review from other reviews in the repository. To match the length of the synthetic review as that of the base review, cosine similarity is used when doing sentence replacement thus preventing review length to be used to detect fake reviews. In order to detect synthetic reviews different coherence measures such as Sentence Transition, Word Co-occurrence and Pairwise sentence similarity measures are used. Sentence Transition is a measure where given a word in a sentence ,one could expect to observe certain words in its following sentence with the same probability. Word Co- occurrence focuses on showing co-occurrence patterns in two consecutive sentences. Pairwise sentence similarity considers the semantic overlap between two consecutive sentences. Here the author has proposed a general framework to detect the synthetic reviews in an efficient manner. The proposed framework has improved the detection accuracy by 13% compared to other deception detection systems. The author here has not focused on the differences in the characteristics of the synthetic reviews from the actual review and needs to identify the parameters of the actual review which differ from the synthetic reviews. Further the proposed review detection mechanisms need to be tested on the parameters of the actual review which differ from the synthetic reviews.

C. Detecting spam review through sentiment analysis

In paper[3] the author detects spam reviews by incorporating the concept of sentiment analysis. He first creates a sentiment lexicon which combines the data present in existing sentiment lexicons such as SentiWordNet and MPQ, along with designing sentiment lexicon specifically for products. Further he calculates sentiment score which is the sentiment polarity of a review .Also he calculates other parameters such as sentiment ratio(ratio of sentiment sentence to all sentences) and difference of sentiment polarity(inconsistency in sentiment score and rating score).Next he constructs discriminative rules to classify the reviews as spam. Finally he combines the discriminative rules with the time series method to detect spam in a spam detection algorithm. Here the reviews are pre-processed and ordered by time. For each subset of reviews ,each review is checked to find whether it satisfies any of the discriminative rules within the given time window.If the result is positive then the store view is categorized as spam. The proposed sentiment score method outperforms the rating and word counting methods with an accuracy of 85.7%.The discriminative rules here are derived based only on three parameters namely sentiment score, sentiment ratio and discrepancy between rating and sentiment. However the author has not considered the effect of behavioural parameters such as Individual rating deviation,Individual content similarity and Individual early time frame for deriving discriminative rules and this is one of the areas which needs to be explored.

D. Neural Network used to detect fake reviews by exploiting product related review features

Detection of fake reviews always depend on labelled data set.In this paper[4] the author proposes a convolutional neural network model which captures the product related review features and a classifier is constructed based on the prod- uct word composition features. The assumption here was review spammers emphasized product features with highly positive or negative words to describe a product. As the product oriented information affects the prediction, integrat- ing this into a classification model is very much essential. The proposed model offers classification results by incor- porating a bagging model. This model combines the affect of 3 classifiers namely PWCC(product word composition classifier),TRIGRAMS-SVM classifier and BIGRAMS-SVM classifier. In PWCC classifier uses product word composition where product and review information are both fed into the classifier for generating predictions.Bigrams-SVM represent each review with bigrams feature set on which an SVM classifier is trained to detect fake reviews. In TRIGRAMS- SVM trigram

feature set is used to build the SVM classi- fier.To get more robust results the proposed bagging model combines the 3 classifiers.It was observed that the bagging model outperforms the other 3 individual models in terms of accuracy. Here the author has not focused more on the reviewer related behavioural features which are likely to make noteworthy contribution to the prediction task.

E. Fraud Detection in Online Reviews By Network Effects

In this paper[5] the author has proposed a model which uses network classification to detect frauds. Here the author speaks about an unsupervised learning method to detect spam reviews. Here the review dataset is represented as a bipartite network where user nodes are connected to product nodes with links representing the review rating. A threshold is set and if the link rating crosses the threshold it is considered positive and negative otherwise. The prior beliefs of users and products can be estimated based on prior information such as review text, timeseries activity and other behavioural features. An iterative algorithm called Signed Inference Algorithm is used which is a message passing algorithm. Here a set of messages are exchanged between the users and products in an iterative fashion until the messages stabilize. Finally the marginal probabilities are calculated and final belief of user having a particular label is computed.The class labels of both users and products are inferred by the final belief vectors. The marginal class probabilities of products, users and reviews are used to order each set of item in a ranked list. The proposed method is very simple and does not require labeled data for classifying reviews. However the author has considered only user ratings as the basis to detect spam and has not considered the sentiment analysis of text found in reviews.

F. Detecting Fake Reviews by the principle of Collective Positive Unlabelled Learning

In this paper[6]the author has proposed a model of detecting fake reviews by learning from positive and unlabelled examples. Here a heterogeneous network consisting of users, reviews and IP addresses is being considered for the learning model where a classification algorithm called MHCC(Multi- typed Heterogeneous Collective Classification)algorithm is used first and then this is extended to Collective Positive and Unlabelled Learning. The MHCC algorithm considers a heterogeneous network of users,IP addresses and reviews as input along with feature matrix of reviews , users and IP addresses .The class label of the review is the desired output. The algorithm has both initialization and prediction steps where the adjacency matrix is computed and then the classifier is trained .The initial classifier gives a rough estimate of the review being in fake review class and the labels of IP and users are derived from majority class labels of their related reviews.

In the prediction step a relation feature matrix is constructed from the estimate labels of the neighbouring nodes and this is used as the basis to train 3 different classifiers for reviews,users and IPs which will provide more accurate results.

However this algorithm treats all unlabelled samples as negative and the adhoc labels of users and IPs may not be accurate as they are derived from labels of neighbouring reviews. Hence an extended model called Collective Positive Unlabelled model is used where new labels are updated with respect to confident positives and negatives from all entity types. This algorithm allows initial labels to be violated if current probability estimate indicates opposite prediction. The model outperforms baseline algorithms like Logistic Regression and also detects a large number of potential fake reviews hidden in an unlabelled set thus improving the training data. Heterogeneous network can be extended to include metadata information of users such as timeframe of writing reviews, number of written reviews and demography of the reviewer .This can be an active area to be explored.

G. Word Order Preserving Convolutional Neural Network used for Spam Detection

In this paper[7] the author has used the word order preserving k-max pooling method in convolutional networks for text classification. Here the author has used a deep learning framework where the deep features of deceptive opinion are summarized to characterize deceptive opinion effectively. Here a 4 layer OPCNN model is proposed consisting of input layer, convolution layer, pooling layer and output layer. The input layer consists of word vector in a sentence followed by a two dimensional matrix. Each opinion statement words are thus expressed in the form of their corresponding word vector. The word vectors on passing through the convolution layer are converted into 1 column feature map. The pooling layer then reduces the output of the convolution layer by using k-max pooling method which selects the maximum value of each feature map. The feature obtained in the pooled layer is then connected and the result is then entered into logistic regression model to assess the probability that the comment is deceptive. Softmax function is used as the regression function. The proposed OPCNN model outperforms other baseline methods such as tf-idf+SVM(term frequency-inverse document frequency+Support Vector

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Machines),Bigram+SVM(Support Vector Machines) and CNN(Convolutional Neural Networks) with an accuracy of 70.10%. However data is artificially annotated and requires more manpower which needs to be improved.

H. Detecting Singleton spam Reviews

In this paper[8] the author proposes two methods to detect singleton review spammers. These spammers are post a single review under a single name and post a large number of fake reviews under different names and they tend to write reviews about the same product by rephrasing the sentences or substituting with synonyms of the previous reviews written. Hence here the author has proposed a semantic similarity method which uses wordnet , one of the largest lexical databases comprising of nouns ,adverbs and adjectives grouped by their synonyms to compute the relatedness between words. This method outperforms the vectorial model in capturing spam reviews. Further the author has proposed another topic based model which aims at extracting product aspects from shorter texts such as user opinions and forums. These topic distributions are then compared with a fixed spam threshold and then classified as spam .The bag of phrases LDA(Latent Dirchlet Allocation) model exhibited more accuracy than the bag of words model. However fixing the threshold is one of the difficult aspect here.

Secondly the number of topics should be known ahead of time and the model also fails to explain topic correlation.

III. COMPARISON OF FAKE REVIEW DETECTION SYSTEMS

There is a lot of research carried out in the field of fake review detection techniques. Each detection mechanism has its own merits and demerits. The below given table summarizes the comparison of the various fake review detection mechanisms techniques.

TABLE I
STUDY OF DIFFERENT FAKE REVIEW DETECTION SYSTEMS

IV. CHALLENGES IN FAKE REVIEW DETECTION

1) Sarcasm in Review text:Sometimes people express sar- casm in review text which may be genuine or fake information related to products. Classification of these reviews into fake and genuine reviews is a difficult task.

2) Implicit Sentiments and contextual Information in Re- view Text: Identifying implicit sentiments and contextual

information in reviews creates a problem in classifying fake reviews as this combines both behavioral analysis and text analysis.

3) Seller Reviewer Collusion: It is found that sometimes sellers fix prizes with the websites promoting their product to enhance the value of the product through fake reviews and sharing the profit with the websites.
These reviews are written by experienced people and it is very difficult to detect these reviews.

4) Domain Dependence Of classifiers: The classifiers trained to detect fake reviews in one domain may not give the same results when used in other domain. Cross domain deception opinion detection is one of the active areas which need to be explored.

5) Getting labelled training dataset:Although there has been a lot of research related to getting labelled dataset for fake review detection, it is still found that getting the right kind of data set is still a major problem and effective methods need to be still explored.

IV. CONCLUSION

This paper discusses the various techniques used for identifying fake reviews. Here an overview of the existing fake review detection methods along with their advantages and limitations are being clearly discussed. Further the current challenges in fake review detection which need to be addressed are also briefly stated in the paper. Fake reviews are growing exponentially along with the enormous amount of online web data. Hence there is a strong need to address the challenges in fake review detection and design suitable algorithms to improve the accuracy in detection. This is one of the active research areas to be explored in the current world

REFERENCES

[1] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in Proceedings of the 21st international conference on World Wide Web. ACM, 2012, pp. 191–200.

[2] H. Sun, A. Morales, and X. Yan, "Synthetic review spamming and de- fense," in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013, pp. 1088–1096.

[3] Q. Peng and M. Zhong, "Detecting spam review through sentiment analysis." JSW, vol. 9, no. 8, pp. 2065–2072, 2014.

[4] C. Sun, Q. Du, and G. Tian, "Exploiting product related review features for fake review detection," Mathematical Problems in Engineering, vol. 2016, 2016.

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

[6] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, "Spotting fake reviews via collective positive-unlabeled learning," in Data Mining (ICDM), 2014 IEEE International Conference on. IEEE, 2014, pp. 899–904.

[7] S. Zhao, Z. Xu, L. Liu, and M. Guo, "Towards accurate deceptive opinion spam detection based on word order-preserving cnn," arXiv preprint arXiv:1711.09181, 2017.

[8] V. Sandulescu and M. Ester, "Detecting singleton review spammers using semantic similarity," in Proceedings of the 24th international conference on World Wide Web. ACM, 2015, pp. 971–976.