Detection of Fake Reviews through Opinion Mining: A Survey

Ashwini M C[1*], M C Padma[2]

[1,2]Dept. Of CSE, PESCE, Mandya, India

Abstract—Opinion mining has played a momentous role in providing product recommendation to users. Efficient recommendation system helps in improving customer satisfaction and also enhances business. The credibility of purchasing a product highly depends on the online reviews. Since not all online reviews are truthful and trustworthy, it is important to develop techniques for detecting review spam, it is possible to conduct review spam detection using various machine learning techniques. We survey the prominent machine learning techniques that have been proposed to solve the problem of review spam detection. This literature survey is done to study the various fake review detection techniques in detail.

Keywords—*Sentiment Analysis; opinion Mining; Fake reviews; Machine learning; Recommendation Systems.*

I. INTRODUCTION

Sentiment analysis (SA) is one of the significant domains of machine learning techniques [1]. Sentiment Analysis (SA) is also known as Opinion Mining (OM), Opinion Mining is one of the greatest dynamic research areas in Natural Language Processing. Sentiment Analysis defines a procedure of mining, classifying, analyzing and describing the feelings or sentiments in the form of word-based data using Machine Learning, Natural Language Processing or Statistics. The two terminologies sentiment analysis or opinion mining are more substitutable. Opinion Mining mine the textual data and evaluates public's attitude around an object whereas sentiment analysis classifies the sentiment articulated in a script then examines it.

There are three main categories in sentiment analysis: document-level, sentence-level, and aspect or Phrase level

Document level

Document level sentiment classification executed on the overall sentiments expressed by authors. Documents classified according to the sentiments instead of topic. It is to summarize the whole document as positive or negative polarity about any object (mobile, car, movie, and politician etc.).

Sentence level

Sentence level sentiment classification models extract the sentences contains opinionated terms, opinion holder and opinionated object. It is one level deep to document level and just concerns to the opinionated words but not the features. Number of positive and negative words counted from sentences if positive words are maximum then opinion about object is positive and if the negative words are more than opinion is negative otherwise neutral.

Aspect or Phrase level

Opinion Mining The phrase level sentiment classification is a much more Pinpointed approach to opinion mining. The phrases that contain opinion words are found out and a phrase level classification is done. But in some other cases, where contextual polarity also matters, the result may not be fully accurate. Negation of words can occur locally. But if there are sentences with negating words which are far apart from the opinion words, phrase level analysis is not desirable. The process is Identifying Opinion Words, the role of negation words and Clauses.

Detection of fake reviews is mainly directed to SA at the document level, A SA technique is applied to classify the documents as real positive and real negative reviews or fake positive and fake negative reviews. Fake negative and fake positive reviews by fraudsters who try to play their competitors existing systems can lead to financial gains for them. This, unfortunately, gives strong incentives to write fake reviews that attempt to intentionally mislead readers by providing unfair reviews to several products for the purpose of damaging their reputation. Detecting such fake reviews is a significant challenge. For example, fake consumer reviews in an e-commerce sector are not only affecting individual consumers but also corrupt purchaser's confidence in online shopping [5].

Some of the typical characteristics of fake reviews are:[11]

1.Less Information about the Reviewer: Users who have fewer 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 timeframe: 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.

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.

In this paper section I presents the introduction to the Sentiment Analysis. Section II contains outline of the related work[11], Section III describes applications areas of opinion mining and sentiment analysis. Section IV outlines the research and challenges of Sentiment Analysis and section V concludes research work with future directions.

II. RELATED WORK

A. Detecting spam review through sentiment analysis

In paper [3] the author detects spam reviews by incorporating the concept of sentiment analysis.

CONTRIBUTION:

1. The author first creates a sentiment lexicon which combines the data present in existing sentiment lexicons such as Senti WordNet and MPQ, along with designing sentiment lexicon specifically for products.
2. Further he calculates sentiment score which is the sentiment polarity of a review.
3. 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).
4. 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.

OBSERVATION:

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 need to be explored.

B. 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.

CONTRIBUTOIN:

1. Here the author speaks about an unsupervised learning method to detect spam reviews.
2. The review dataset is represented as a bipartite network where user nodes are connected to product nodes with links representing the review rating.
3. A threshold is set and if the link rating crosses the threshold it is considered positive and negative otherwise.
4. The prior beliefs of users and products can be estimated based on prior information such as review text, timeseries activity and other behavioural features.
5. An iterative algorithm called Signed Inference Algorithm is used which is a message passing algorithm.
6. Here a set of messages are exchanged between the users and products in an iterative fashion until the messages stabilize.
7. 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 items in a ranked list.

OBSERVATION:

The proposed method is very simple and does not require labelled 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.

C. Detecting Fake Reviews by the principle of Collective Positive Unlabeled Learning

In this paper [6] the author has proposed a model of detecting fake reviews by learning from positive and unlabelled examples.

CONTRIBUTION:

1. 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 Unlabeled Learning.
2. The MHCC algorithm considers a heterogeneous network of users IP addresses and reviews as input along with feature matrix of reviews, user sand IP addresses.
3. 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.

OBSERVATION:

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 train3 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 unlabeled 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.

D. 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 Paper [1] the author uses a frequent itemset mining method to find a set of candidate groups.

CONTRIBUTION:

1. Here frequent itemset refers to the set of multiple products reviewed by the users together.
2. Further a minimum support count of three is set which indicates each group should have worked together on at least three products together.
3. Next the degree of spam city 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.
4. This was used as the basis for ranking the spam groups.
5. 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.
6. Finally, an iterative ranking algorithm called GS Rank (Group Spam Ranking Algorithm) is executed to draw inferences from the three relational models and classify spam reviews.

OBSERVATION:

This algorithm outperforms all baseline methods like regression, rank and classification algorithms. However here the author has focused only on the behavioural 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.

E. 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.

CONTRIBUTION:

1. To match the length of the synthetic review as that of the base review, cosine
2. Similarity is used when doing sentence replacement thus preventing review length to be used to detect fake reviews.
3. In order to detect synthetic reviews different coherence measures such as Sentence Transition,

Word Co-occurrence and Pairwise sentence similarity measures are used.

4. 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.
5. Pairwise sentence similarity considers the semantic overlap between two consecutive sentences.

OBSERVATION:

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.

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

Detection of fake reviews always depend on labelled dataset.
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 product word composition features.

CONTRIBUTION:

1. The assumption here was review spammers emphasized product features with highly positive or negative words to describe a product.
2. As the product-oriented information affects the prediction, integrating this into a classification model is very much essential.

OBSERVATION:

The proposed model offers classification results by incorporating a bagging model.
This model combines the effect of 3 classifiers namely PWCC (product word composition classifier), TRIGRAMS-SVM classifier and BIGRAMSSVM classifier. In PWCC classifier uses product word composition where product and review information are both fed into the classifier for generating predictions.
Bigrams-SVM represents 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 classifier. 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.

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.

CONTRIBUTION:

1. Here the author has used a deep learning framework where the deep features of deceptive opinion are summarized to characterize deceptive opinion effectively.
2. Here a 4-layer OPCNN model is proposed consisting of input layer, convolution layer, pooling layer and output layer.
3. The input layer consists of word vector in a sentence followed by a two-dimensional matrix.
4. Each opinion statement words are thus expressed in the form of their corresponding word vector.
5. The word vectors on passing through the convolution layer are converted into 1 column feature map.
6. 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.
7. 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.
8. SoftMax function is used as the regression function.

OBSERVATION:

The proposed OPCNN model outperforms other baseline methods such as tf-idf+SVM (term frequency-inverse document frequency Support Vector
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.

CONTRIBUTION:

1. Hence here the author has proposed a semantic similarity method which uses wordnet, one of the largest lexical databases comprising of noun, adverbs and adjectives grouped by their synonyms to compute the relatedness between words.
2. This method outperforms the vectorial model in capturing spam reviews.

OBSERVATION:

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 Dirichlet Allocation) model exhibited more accuracy than the bag of words model. However, fixing the threshold is one of the difficult aspects here. The number of topics should be known ahead of time and the model also fails to explain topic correlation.

III. APPLICATIONS AREAS OF OPINION MINING AND SENTIMENT ANALYSIS

Sentiment analysis has huge quantity of applications in the Natural Language Processing field. The natural language processing community shows much interest in Sentiment Analysis and Opinion Mining system. The explosion of internet has changed the people's life style, now they are more expressive on their views and opinions [1], and this tendency helped the researchers in getting user-generated content easily.
The major applications of Opinion mining and sentiment analysis are specified as below:

1. Word of Mouth (WOM)

It is human nature to judge a metaphorical book by its reviews. No matter how good a product is or how good a user's experience with it is, they tend to believe in the word of mouth (WOM). Thus, Fake reviews are a hidden threat to e-commerce businesses. It is unethical but widely practiced. Consequently, this is going to be where Sentiment Analysis comes into depiction. As the online appraisal blogs, sites, social networking sites have provided the great amount of opinions; it has facilitated in the development of decision making so a lot easier for the user.

2. Voice of Voters

Each of the political parties typically depleted a major portion of the quantity of money for intend of demonstration for their party or for persuade the voters. Accordingly, if the politicians recognize the people opinions, reviews, implications, these can be completed with more outcomes. This is how procedure of Sentimental analysis does not only assist political parties but on the other hand help the news analysts beside. Also, the British and the American management had previously used a few of the similar practice.

3. Online Commerce

There is vast number of websites related to ecommerce. Majority of them had the policy of getting the feedback from its users and customers. After getting information from various areas like service and quality details of the users of company users experience about features, product and any suggestions. These details and reviews have been collected by company and conversion of data into the geographical form with the updates of the recent online commerce websites who use these current techniques.

4. Purchasing Product or Service

While purchasing a product or service, taking right decision is no longer a difficult task. By this technique, people can easily evaluate other's opinion and experience about any product or service and also he can easily compare the competing brands. Now people don't want to rely on external consultant. The Opinion mining and sentiment analysis extract people opinion form the huge collection of unstructured content, the internet, and analyze it and then present to them in highly structured and understandable manner.

5. Quality Improvement in Product or service

By Opinion mining and sentiment analysis the manufactures can collect the critic's opinion as well as the favourable opinion about their product or service and thereby they can improve the quality of their product or service. They can make use of online product reviews from websites such as Amazon and C|Net [2, 3], RottenTomatoes.com [4] and IMDb [5].

6. Voice of the Market

At any time, a product is to be commenced by a definite company, the customers would to recognize about the product ratings, reviews and comprehensive metaphors about it. Sentiment Analysis can assist in analyzing marketing, advertising and for making new tactics for endorse the product. It offers the customer a chance to prefer the best among the all.

7. Brand Reputation Management

Sentiment analysis would help to decide how would be a company's brand, service and the service or product that would be alleged by the online society. Brand Reputation Management will be worried about the administration of the reputation of market. It has spotlight on the company and product moderately than customer. Thus, the prospects were formed for the rationale of managing and escalating the brand reputation of the organizations.

8. Government

Sentiment Analysis has facilitated the administration for the purpose of offering a variety of services to the public. Fair outcomes have to be produced for analyzing the negative and positive points of government. Thus, sentiment analysis is helpful in many areas like decision making policies, recruitments, taxation and evaluates social strategies. Some of the parallel techniques that offer the citizen-oriented government model where the services and the main concern should be presented as per the citizens. One of the attractive problems which can be taken up is applying this method in the multilingual country like the India where content of the generating mixture of the different languages (e.g. Bengali English) is a very common practice.

9. Recommendation Systems

By classifying the people's opinion into positive and negative, the system can say which one should get recommended and which one should not get recommended [8].

10. Detection of "flame"

The monitoring of newsgroup and forums, blogs and social media is easily possible by sentiment analysis. Opinion mining and sentiment analysis can automatically detect arrogant words [6], over heated words or hatred language used in emails or forum entries or tweets on various internet sources.

11. Opinion spam detection

Since internet is available to all, anyone can put anything on internet, this amplified the possibility of spam content on the web. People may write spam content to mislead the people. Opinion mining and sentiment analysis can classify the internet content into" spam" content and „not spam" content [1].

12. Policy Making

Through Sentiment analysis, policy makers can take citizen's point of view towards some policy and they can utilize this information in creating new citizen friendly policy.

13. Decision Making

People's opinion and experience are very useful element in decision making process. Opinion mining and Sentiment analysis gives analyzed people's opinion that can be effectively used for decision making.

IV. RESEARCH CHALLENGES IN OPINION MINING AND SENTIMENT ANALYSIS

1) Detection of spam and fake reviews: The web holds both genuine and spam contents. For efficient Sentiment classification, this spam content is supposed to be eliminated before processing. This can be done by discovering duplicates, by perceiving outliers and by bearing in mind reputation of reviewer.

2) Limitation of classification filtering: There is a restriction in classification filtering while shaping most trendy thought or idea. For enhanced sentiment classification result this restriction should be condensed. The risk of filter bubble [11] gives immaterial opinion sets and its outcome is false accumulation of sentiment.

3) Asymmetry in availability of opinion mining software: The opinion mining software is very costly and at present reasonable only to big organizations and government. It is beyond the common citizen's anticipation. This should be accessible to all people, so that everybody acquires profit from it.

4) Incorporation of opinion with implicit and behaviour data: For flourishing examination of sentiment, the opinion words should assimilate with implied data. The inherent data decide the actual conduct of sentiment words.

5) Domain-independence: The major confronts faced by opinion mining and sentiment analysis is the area reliant scenery of sentiment words. One feature set may provide high-quality performance in one field, at the same time it carries out very poor in some other field.

6) Natural language processing overheads: The natural language above your head like uncertainty, co-reference, obliqueness, conclusion etc. created difficulty in sentiment analysis too.

7) 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.

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

9) Sarcasm in Review text: Sometimes people express sarcasm 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.

10) 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 behavioural analysis and text analysis.

11) 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.

V. CONCLUSION AND FUTURE SCOPE

This paper mainly focuses on survey of opinion mining. Opinion mining or sentiment analysis has a surplus amount of applications in the information systems, which includes classification of reviews, their summarization and a variety of real-time applications. Researches have been conducted to mine the opinions in the form of document level, sentence level, aspect or phrase level sentiment analysis.

Thus, we have discussed about the overview of challenges in opinion mining and the various applications of sentiment analysis. More future research can be done on these challenges and more work can be done to overcome these challenges in future.

REFERENCES

[1] A. uike, W. Ahmed *"Improved Text Minig Techniques for spam Review Detection"*, IJCSE,vol.7,issue.5,pp.147-152- ,2019.

[2] Pooja, k. Bhatia *"Spam detection using Naïve Bayes Classifier"*, IJCSE , vol.6, issue.7, pp.934-938.

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

[6] 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 pp. 191–200,2012.

[7] 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, pp. 1088–1096,2013.

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

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

[10] V. Sandules cu 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.

[11] A .Lakshmi Holla , Dr Kavitha K.S "A Comparative study on fake review Detection Techniques", IJERCSE,vol.5,issue.4,pp.641645,2018.