DETECTION OF REVIEW SPAM AND REVIEW SPAMMERS GROUP

CH. RATHAN KUMAR[1], Dr. K. RADHIKA[2]
[1]*Research Scholar, CSE Department, OSMANIA UNIVERSITY, Hyderabad, India*
[2] *Supervisor, Professor, IT Department, CBIT, Hyderabad, India*

***Abstract*—** In the past few years, online reviews are hugely popular and crucial resource of customers' opinions. These reviews are useful for individual people to buy products and business organizations to take business decisions. But for the purpose of profits and to gain popularity some spam reviews will be given by fraudulent people. The fraudulent activities misinform certain customers and organizations remolding their businesses and forbid opinion-mining techniques from reaching exact conclusions. To detect the spam reviews, the recent time researches concentrate on systematically examining and also categorizing the models for detecting the spam reviews. In this paper, in order to solve spam reviews problem we will study some machine learning techniques that have been proposed and we will study the performance of different approaches for classification and detection of review spam. This paper major goal is to provide a solid and comparative study of today's research on detecting spam reviews and review spammers group using different machine learning techniques and Comparative summary of detecting spam reviews, and spam reviewers group detection techniques.

***Keywords*—** Supervised learning, unsupervised learning

I. INTRODUCTION

As the Internet continues to grow in both size and importance, the quantity and impact of online reviews continually increases. Reviews have power to influence individuals across abroad range of industries, but are majorly important in the region of ecommerce, where comments and reviews related to products and also services are the highly convenient, if not the only, way for a purchaser to take a decision on whether or not to purchase them. Online reviews are written for a variety of reasons. Often, in an effort to improve and enhance their businesses, online retailers and service providers sometimes request their customers to give feedback about their experience regarding the products or services they have purchased and asks if they were satisfied about the product or not. Customers may also feel that it's better to give review on a product or service if they had a good or bad or worst experience with the product.

While reviews on online can be helpful, but blind trust of these reviews is dangerous for both the seller and buyer. A lot of people wanted to read online reviews before placing any order on online. The reviews may be false hyped or faked for profit or gain. Hence we have to be careful before we take any decision by reading online reviews.

Furthermore, business owners will give money to the persons who writes good reviews about their own commodities and also they encourage the reviewers to write bad reviews about others products or services. These reviews are considered spam reviews. These reviews can have a huge impact in online marketing areas.

In 2011 NirKshetri et al. [1] discovered the illegal and unethical practices and cybercrimes in social media. Then in 2013 Marco Huesch ,Greg VerSteeg et al.[2] identified vulnerabilities in social media content and they explored Manipulation of public opinion identified to detect bad practices over social media.

Similarly in 2017 Summer Lightfoot et al [3] provides useful insight about false propaganda (fake reviews) over social networks.

II. KINDS OF SPAM REVIEW (OPINION SPAM)

According to Dixit et al. [4] opinion spams categorized into three groups. They are Untruthful reviews, Reviews on brands only, Non-reviews.

Untruthful reviews: Also called fake or bogus reviews, these are very virulent and their purpose is to intentionally misguide readers or customers or automated systems by reading false positive or false negative reviews about a product or service.

Reviews on brands only: These reviews do not contain specific product or service reviews but these reviews for brands, manufacturers, or sellers.

Non-reviews: These are not actual reviews or opinions. They may be advertisement or other irrelevant text which contain no opinion.

III. FEATURE ENGINEERING FOR DETECTING SPAM REVIEWS

It is important to specify that while a lot of existing

techniques of machine learning are not enough effective for detecting spam reviews, they can have been discovered to be more reliable than manual detection. These issues are identified by Abbasi et al. [5], is the lack of any differentiating words or features that gave a definitive clue for classification of reviews as actual or fake. A general approach in text mining is to use a set of words approach where the presence of word, or small unit of words are used as features. Many of the studies found that the above mentioned approach is not sufficient to train a classifier with enough performance in detecting spam reviews.

Therefore, additional methods of feature engineering must be examined in an effort to extract an additional informative feature set that will improve spam review detection. Types of Feature's used in detection of spam reviews:

1) Linguistic features (or) Review centric features: - Review centric features are features that are constructed using information comprised in a single review. Categories in this feature are:

i) Bag of words: In a bag of words approach, individual or small groups of words from the text are used as features, are called as n-grams. These are made by choosing n continuous words from a given sequence. That means selecting 1, 2 or 3 contiguous words from a text. They are denoted as a uni-gram, bi-gram, and tri-gram (n = 1, 2 and 3) respectively.

ii) Term frequency: These features are similar to bag of words however, instead of simply being concerned with the existence or non-existence of a term; it concerns the frequency with which a term occurs in each review, so we include the count of occurrences of a term in the review.

iii) Part of Speech (POS) tagging: It involves tagging word features with a part of speech based on the definition and its circumstances within the sentence in which it is found.

iv) Word Count (WC): It is a text analysis software tool in which users can build their own dictionaries to study dimensions of language especially their points of interest.

v) Stylometric features: These features are either character and word based lexical features or syntactic features. Lexical features gives suggestion of the types of words and characters that the writer wishes to use and includes features such as average word length or the number of upper case characters. Syntactic features try to represent the reviewer's writing style and include features such as the amount of punctuation words such as "a", "the", and "of".

vi) Semantic features: These features address the underlying meaning of the words used to make semantic language models for detecting fake reviews.

vii) Review characteristic or metadata: These features contain metadata (information about the reviews) rather than information on the text content of the review. These characteristics could be the review's length, date, time, rating, reviewer id, review id, store id or feedback.

2) Behavioral features (or) Reviewer centric features: - These will take a holistic look at all of the reviews written by any particular author, along with information about the particular author.

i) Maximum number of reviews: It was found that about 75 % of spammers write more than 5 reviews on any given day. Therefore, taking into account the number of reviews a user writes per day can help detect spammers.

ii) Percentage of positive reviews: Approximately 85 % of spammers wrote more than 80 % of their reviews as positive reviews, thus a high percentage of positive reviews might be an indication of an untrustworthy reviewer.

iii) Review length: The average review length is a very significant aspect of reviewers with suspicious intentions since about 80 % of spammers won't write reviews more than 135 words.

iv) Reviewer deviation: It was analyzed that ratings of spammers tend to vary from the average review rating at a far higher than the rate of legitimate reviewers, therefore identifying user rating variations might be helpful in detection of dishonest reviewers.

v) Maximum content similarity: The presence of like reviews for many different products or goods by the same reviewer has been known to be an indication of a spammer.

3) Information about the product: - Information about a product is useful in spam detection such as, the product description and sales volume, information about merchandise being reviewed as average ratings, number of reviews, product description, popularity and sales volume. For example, a product with many positive reviews but low sales calls the reliability of the positive reviews into question.

IV. SPAM OR FAKE REVIEW DETECTION USING MACHINE LEARNING TECHNIQUES

In this paper we discuss machine learning techniques that have been proposed for the detection of online spam review with an emphasis on feature engineering. The identification of opinion spam has become a huge concern

in today's times to authenticate online reviews and gain consumer faith, trust and confidence.

Detection types: -

Review centric spam review detection.
Reviewer centric review spam detection.

1) Review centric review spam detection: -

It is the most usual form of review spam detection, which uses machine learning techniques to develop models using the content and metadata of the reviews. Supervised learning is the task of learning from labelled data and it is the most frequent method used for review spam detection in the literature. This method requires labeled information or data in order to train a classifier, on the other hand, unsupervised learning uses unlabeled data to find unseen relationships between instances independent of a class attribute, Semi-supervised learning is a combination of both supervised and unsupervised learning, which uses a few labeled instances in combination with a large number of unlabeled instances to train a classifier.

1) Supervised learning: - Supervised learning can be used to find review spam by looking at it as the classification problem of separating reviews into two classes: spam and non-spam reviews. Initially Jindal et al [6] discussed the progression of opinion mining, they found that opinion spam is totally different from email and Web spams.

He primarily focused on summarizing extracting or the opinions from text by using Natural Language Processing (NLP). Next Jindal et al collected millions of reviews on products from amazon, categorized reviews and identified spam reviews using near duplicate reviews method. Raymond et al. [7] identified another set of features from reviews and used logistic regression model to identify fake reviews. Raymond et al. got AUC score of 0.78 was achieved when using all features, compared to an AUC score of 0.63 when only using text features. Ottet al.[8] produced dataset using Amazon Mechanical Turk (AMT) in combination with Trip Advisor.

For this work, three groups of features were identified: POS tag frequencies, WC, and bigram for text categorization based features. Naïve Bayes and SVM classifiers were trained and evaluated, their best model achieved an accuracy of 89.8 % using bigram and WC features with an SVM classifier. Li et al.[9] created a cross domain dataset that included three types of reviews from three domains (hotel, restaurant and doctor).

His classification framework was based on using the Sparse Additive Generative Model (SAGE), which is a generative Bayesian approach. Shojaee et al.[10] proposed a novel method for detecting review spam by using Stylometric (Lexical and Syntactic) features. In this work they developed classifiers on the dataset created by Ott et al.They observed that the hybrid feature set using the SVM learner achieved the highest performance, an F-measure of 84 %.

2. Unsupervised Learning: - The use of supervised learning method are not applicable Because of the difficulty of constructing accurately labeled datasets of review spam. It provides a solution for this because it doesn't need labeled data.

A novel unsupervised text mining models are developed and combined into a semantic language model for identifying false reviews by Raymond et al.[7] and this work compared with supervised learning methods.

An unsupervised method proposed by Wu et al. (2010) [11] shows the effect of distortion in distinguishing positive singleton spam reviews from positive singleton real reviews on a dataset of hotel reviews.

A novel generative model called Latent Spam Model (LSM) [2014] [12] for spam review detection using unsupervised learning developed by Arjun Mukherjee et.

3) Semi-supervised learning: - In other domains, it has been discovered that utilizing unlabeled data in addition with a little amount of labeled data can gradually improve learner accuracy as compared to completely supervised methods. In a study by Li et al.,[13] a two-view semi supervised method for review spam detection was created by employing the framework of a co-training algorithm to make use of the large amount of unlabeled reviews available. PU-Learning is another type of semi-supervised learning approach, developed by Liu et al.The model is prepared and evaluated utilizing all of the unlabeled data as the negative class and any instances that are classified as positive are removed.

2) Reviewer centric review spam detection:-

Identifying reviewers who are creating fake reviews are given importance in the effort to detect spam reviews. Using reviewer centric features in collaboration with review centric features might be chosen over a review centric only approach for detecting spams.

Additionally, collecting behavioral proofs of spammers is easier than recognizing spam reviews. Mukherjee et al. study of supervised learning approaches for deceptive review detection observed that using behavioral features yields higher performance than linguistic features alone on the real world Yelp dataset. Behavioral features (i.e., higher percentage of positive reviews, high number of reviews, average review length).

V. GROUP SPAM REVIEWERS DETECTION

Occasionally, spamming activities can be considered the events of group spamming; manufacturers hires more spammers to do a task because they can have ability to dominate all aspects, features and opinions for a product or brand. On another times, the persons will work together geographically and they are in contact with each other. This process will increase their abilities, power and cooperation at the time of attacks.

Various behaviors of spamming can be extracted from groups of spammers. These are used to classify spam groups from individual reviewers. The features used in group spam detection in Mukherjee et al. (2012) [14] are called spam indicators.

Features used to find group spammers:

1. Number of reviews within a time interval
2. Deviations between the average ratings of a product and the ratings given by members of the group.
3. Content similarity between members;
4. Content similarity among a group;
5. Group early time frame
6. Group size.
7. Group size ratio,
8. Group support count:
9. Individual Member Coupling

In the previous works of group spammers detection Mukherjee et al. [2012] proposed GRank as a relational model used as relationships between individual and group indicators and target products to rank candidate groups as spam or non-spam groups using supervised learning.

Zhuo Wang [2015] [15] proposed a Review Spammer Groups via Bipartite Graph Projection, which is loose spammer group detection problem and he obtained good precision and recall compared to frequent item set mining (FIM) FIM-based approach. LU ZHANG [2017] [16] propose a partially supervised learning model (PSGD) to detect spammer groups.

PSGD applied PU-Learning to study a classier as spammer group detector from positive instances (labeled spammer groups) and unlabeled instances (unlabeled groups). Experiments on Amazon.cn data set shows that the proposed method is effective compared to the state-of-the-art spammer group detection methods (Naive Bayesian model and an EM algorithm).

VI) COMPARATIVE SUMMARY OF REVIEW SPAM DETECTION, SPAM REVIEWERS GROUP DETECTION TECHNIQUES

Table 6.1: Summary for spam review detection techniques

Author	Title	Year	Journal	Data set(s) used	Performance Metric
Kyumin Lee, James	Detecting Collective Attention Spam	2012	ACM	Twitter dataset	Accuracy, false positive rate and false negative rate and total spam detection
Xia Hu, Jiliang Tang,	Social Spammer Detection with Sentiment Information	2014	IEEE	Twitter dataset	Precision, recall, and F1-measure
Kristin Kinmont	Fake News Detection in Twitter	2014	IEEE	Twitter data	Truthy, TweetCred and Cognos.
Yuqing Lu, Lei Zhang	Simultaneously Detecting Fake Reviews and Review Spammers using Factor Graph Model	2013	ACM	Amazon Dataset	average F1 and Accuracy
Arjun Mukherjee Vivek Venkataramn	Opinion Spam Detection: An Unsupervised Approach using Generative Models	2014	Semantic Scholar	AMT Dataset, Amazon Dataset, Yelp Restaurants	precision, recall, and F1-score
Shebuti Rayana Leman Akoglu	Collective Opinion Spam Detection: Bridging Review Networks and Metadata	2015	ACM	Yelp.com	AP and AUC
JITENDRA KUMAR ROUT	Revisiting Semi-Supervised Learning for Online Deceptive Review Detection	2017	IEEE	Gold standard dataset by Ott et al.	Accuracy, precision, Recall, F-score

Saeedreza Shehnepoor, Mostafa	NetSpam: a Network-based Spam Detection Framework for Reviews in Online Social Media	2017	IEEE JOURNAL	Yelp dataset	AP and AUC
Dilsha, lijo	Opinion Spam Setection using Review, Reviewer Centric features	2017	IEEE	food product data set	F-Score
Man-Chun Ko,	Paid Review and Paid Writer Detection	2017	ACM	restaurant reviews from Pixnet	Precision, Recall, F1
Huaxun Deng, Linfeng Zhao	Semi-supervised Learning based Fake Review Detection	2017	IEEE	crawled from JD.com	Accuracy
Wael Etaiwi, Arafat Awajan	The Effects of Features Selection Methods on Spam Review Detection Performance	2017	IEEE	`gold standard' dataset by Ott et al.	Precision, Recall, Accuracy
Simran Bajaj, Niharika Garg	A Novel User-based Spam Review Detection	2017	Elsevier	own dataset	Accuracy
Draško Radovanovi	Review Spam Detection using Machine Learning	2018	IEEE	Akismet	Accuracy
Arjun Mukherjee, Bing Liu	Detecting Group Review Spam	2011	ACM	Amazon Dataset	Avg Number of detected spam groups
Arjun Mukherjee Bing Liu, Natalie Glance	Spotting Fake Reviewer Groups in Consumer Reviews	2012	ACM	Amazon Dataset	AUC (Area Under the ROC Curve)
Zhuo Wang, Tingting Hou, Dawei Song	Detecting Review Spammer Groups via Bipartite Graph Projection	2015	British Computer Society	amazon review dataset	Precision, recall and F1, Number of k-connectivity spam groups
Zhuo Wang · Songmin Gu · Xiangnan Zhao	Graph-based review spammer group detection	2017	Springer	amezon yelp.com	Precision, Recall and F1-Score
LU ZHANG , ZHIANG WU,	Detecting Spammer Groups From Product Reviews: A Partially Supervised Learning Model	2017	IEEE	Amazon.cn	Precision, Recall and F1-Score

VII. CONCLUSION

To understand the trends for detecting spam reviews and future directions for researchers on review spam detection, in our study we provided different types of features and two main approaches used for review and reviewers spam detection. Along with them this survey provided metrics used to find group spammers in opinion spam reviewer's detection which is a broad future work in this area. This survey also provided summary table for spam review and reviewers detection which contains previous work done by researchers in this area and provided their performance metrics. As per detected gaps in literature survey, future work will be extracting the most effective features from reviews and reviewers to find spam reviews using unsupervised learning method which uses unlabeled data or raw data and to increase the accuracy of detection because most of the previous works are developed on supervised method on labeled data for detection.

REFERENCES

[1] Nir Kshetri Privacy and Security Aspects of Social Media: Institutional and Technological Environment Pacific Asia Journal of the Association for Information Systems Vol. 3 No. 4, pp.1-20 / December 2011
[2] Marco Huesch Greg Ver Steeg and Aram Galstyan Vaccination (Anti-) Campaigns in Social Media Expanding the Boundaries of Health Informatics Using Artificial Intelligence: Papers from the AAAI 2013 Workshop
[3] Summer Lightfoot Political Propaganda Spread through Social Bots (2017) https://www.researchgate.net/publication/324024528

[4] Dixit S, Agrawal AJ (2013) Survey on review spam detection. Int J Comput Commun Technol ISSN (PRINT) 4:0975–7449
[5] Abbasi A, Zhang Z, Zimbra D, Chen H, Nunamaker JF Jr (2010) Detecting fake websites: the contribution of statistical learning theory. MIS Q 34(3):435–461
[6] Jindal N, Liu B (2008) Opinion spam and analysis. In: Proceedings of the 2008 International Conference on Web Search and Data Mining (pp. 219–230). ACM, Stanford, CA
[7] Lau RY, Liao SY, Kwok RCW, Xu K, Xia Y, Li Y (2011) Text mining and probabilistic language modeling for online review spam detecting. ACM Trans Manage Inf Syst 2(4):1–30
[8] Ott M, Choi Y, Cardie C, Hancock JT (2011) 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).
[9] Li J, Ott M, Cardie C, Hovy E (2014) Towards a general rule for identifying deceptive opinion spam. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 1566–1576, Baltimore, Maryland, USA, June 23-25 2014. ACL
[10] Shojaee S, Murad MAA, Bin Azman A, Sharef NM, Nadali S (2013) Detecting deceptive reviews using lexical and syntactic features. In: Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on (pp. 53–58). IEEE, Serdang, Malaysia
[11] Wu et al. (2010). Distortion as a validation criterion in the identification of suspicious reviews. Social media analytics (pp. 4).
[12] Arjun Mukherjee (2014) Opinion Spam Detection: An Unsupervised Approach using Generative Models Semantic Scholar.
[13] Li F, Huang M, Yang Y, Zhu X (2011) Learning to identify review spam. In: IJCAI Proceedings-International Joint Conference on Artificial Intelligence, vol 22, No. 3., p 2488
[14] Mukherjee A, Liu B, Glance N (2012) Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st international conference on World Wide Web. (pp. 191–200). ACM, Lyon, France.
[15] Zhuo Wang∗, Tingting Hou, Dawei Song, Zhun Li and Tianqi Kong Detecting Review Spammer Groups via Bipartite Graph Projection Computational Intelligence, Machine Learning and Data Analytics The Computer Journal, 2015.
[16] LU ZHANG , ZHIANG WU, (Member, IEEE), AND JIE CAO Detecting Spammer Groups From Product Reviews: A Partially Supervised Learning Model VOLUME 6, 2018 2169-3536 2017 IEEE.