A REVIEW ON FAKE PRODUCT REVIEW DETECTION AND REMOVAL TECHNIQUES

[1] RUTUJA B. ARDAK, [2] PROF. GIRISH S. THAKARE

[1] P.G Student, Department of Computer Science & Engineering, Sipna C.O.E.T, Amravati, India,
[2] Associate Professor, Department of Computer Science & Engineering, Sipna C.O.E.T, Amravati, India.

Abstract:

As we can see within the last couple of decades, the trend of Online Shopping has become a convenient option for us all. Due to the growing need and demand of Online Shopping and Online Businesses, the businessmen need to constantly rely on Computer Science and technologies to grasp the requirements of the customer. However, due to the scope of consumers growing rapidly, and on a wider scale on the internet, it becomes troublesome for interested personnel to actually grasp the necessary reviews to evaluate a product. Moreover, there are always some fake reviews of the product. This provides difficulties, both to the customer, and to the product manufacturers who try to consider the customer requirements. As e-commerce is growing and becoming popular day-by-day, the number of reviews received from customer about the product grows rapidly. For a popular product, the reviews can go upto thousands. This creates difficulty for the potential customer to read them and to make a decision whether to buy or not the product. This paper implements a method that utilizes data mining to work on genuine product reviews by genuine customers, lets the manufacturers and other customers know whether the review is positive or negative, and also blacklists fake accounts. The user will first have to log in using their email accounts for verification, then they would be given a summary of reviews based on desired products and different features of the product, then after their own reviewing, the comment will automatically be checked whether it's a positive or a negative one. Similarly, fake accounts will get blacklisted and will be denied further access. After all this, the reviews would be updated for further information to be viewed. This technique for review monitoring proved to be effective and efficient based on the experimental analysis and surveys.

Index Terms – **Online product review, Product review, Online Fake/Spam reviews, Fake review detection techniques.**

I. INTRODUCTION

Ever since the beginning of modernization of humanity, buying desired commodities and goods have been continuing forever. We often recommend our friends, relatives or acquaintances what products to buy, and what not based on personal experience. Similarly, when we want to buy something we never have, we consult those people who have some experience in that area. This method of review gathering has also been essential for manufacturers to decide what products or product features should best satisfy the customers.

This trend has since continued, but it has taken the form of online reviews in today's digital age. Due to the emerging of E-Commerce in these times, it has become necessary to focus on such online elements of business. Fortunately, almost all of the E-commerce companies have implemented the usage of review systems along with their products. However, in accordance with such a large number of online connected people, residing in different parts of the globe, maintaining the reviews and organization of them becomes troublesome. To overcome these challenges, we have introduced a new Review monitoring system through this paper. This system will help organize the user reviews so that both the potential customers as well as the manufacturers can easily make decisions regarding buying or selling of different products.

Our method will provide the users with the facility to find reviews on a particular product, as well as on specific features of a product. To give them a clearer picture, this system will not only mine the reviews information, but also categorize the reviews on the basis of positive or negative reviews. This helps the users to make decisions based on the plus points as well as the faults encountered by previous customers. As for the manufacturers, it will help them receive the feedback that users have given to their product, based on which they can decide whether to continue with the product, or to make the necessary improvements. Also, due to the online platforms, there are chances that some the reviews are false or inappropriate. Our system also addresses this

problem as there will be a proper user verification, and irrelevant or suspicious accounts will get banned, following which they won't be given access to that particular website. This serves as a great system for filtering the reviews and maintain a nice interface which pleases the users.

II. LITERATURE REVIEW

[1] In the year 2011, Review Graph Based Online Store Review Spammer Detection was proposed. In this paper, researchers proposed a novel concept of a heterogeneous review graph to capture the relationships among reviewers, reviews and stores that the reviewers have reviewed. They explored how interactions between nodes in this graph can reveal the cause of spam and propose an iterative model to identify suspicious reviewers. This is the first time such intricate relationships have been identified for review spam detection. They also develop an effective computation method to quantify the trustiness of reviewers, the honesty of reviews, and the reliability of stores. Different from existing approaches, they didn't used review text information. The model is thus complementary to existing approaches and able to find more difficult and subtle spamming activities, which are agreed upon by human judges after they evaluate the results.

[2] In 2014, there was a paper published on Spotting Fake Reviews via Collective Positive - Unlabeled Learning. In this paper researchers have studied Dianping the largest Chinese review hosting site and also filtered reviews from Dianping's fake review detection system .Dianping's algorithm has a very high precision but the recall is hard to know. This paper studied the problem of fake review detection in the Collective PU learning framework. This paper first proposes supervised learning algorithm MHCC for the heterogenous network of reviews, users and IPs and then extended it to CPU model which is more appropriate for PU learning problem because the labels of reviews have very high precision but unknown recall. With the labeled data provided by the review hosting site Dianping, they have conducted several experiments to show that combining collective classification and PU learning ,proposed CPU model has major advantages over existing state of the art baseline algorithm. This model has language independent features because of which model is generalized to many other languages.

[3] In the year 2015, the paper named as BILD Testing for Spotting out Suspicious Reviews, Suspicious Reviewers, and Group Spammers was introduced. In this paper, researchers expects to present parameterized methodology for distinguishing suspicions by making use of metadata about reviews. , Online audits give profitable data about items and administrations to buyers. On the other hand, spammers are joining the group attempting to deceive pursuers by composing fake surveys. Past endeavors for spammer identification utilized reviewers' behaviors, text similarity, linguistics features and rating patterns. Those studies have the capacity recognize certain sorts of spammers, e.g., the individuals who post numerous comparable surveys about one target substance. s. Metadata about reviews is private data like MAC address IP Address, location, date- time and browser id. In this way, BILD TEST method to identifies suspicious reviews, reviewers and group spammers on shopping websites.

[4] In 2017, a paper was published on RLOSD: Representation Learning Based Opinion Spam Detection. In this paper, the researchers have proposed a decision tree-based method to reveal deceptive reviews from trustworthy ones. They had used unsupervised representation learning along with traditional feature selection methods to extract appropriate features and evaluate them with a decision tree. The model takes data correlation into consideration to opt suitable features. A combinational model composed of feature engineering and feature reduction phases is built to reduce the dimensionality of the feature space and eliminate redundant and irrelevant features which improve the performance of review spam detection. In the first step, a pre-processing process is done. This procedure is done by removing stopwords, stemming and POS tagging. In the next level, feature engineering techniques, e.g. finally, for distinguish between spam and non-spam reviews a classifier is employed in which using Information Gain to rank features and detect deceptive reviews. The result shows the better performance in detecting opinion spam, comparing most common methods in this area.

[5] In 2017, a research paper was proposed which was titled as Opinion Mining using Ontological Spam Detection. In this paper, the proposed technique includes Ontology, Geo location, and IP address tracking, Spam words Dictionary using Naïve Bayes, Brand only review detection and tracking account used to detect Spam reviews in order to get more accurate results from Opinion mining. In current scenario, the data on the web is growing exponentially. As there are a number of fake product reviews, so opinion mining technique incorporates spam detection to produce a genuine opinion. The online shopping is mainly influenced by the reviews but all these reviews are might not real or trustworthy, some reviews are manipulated by suppliers, vendors or publishers to promote their sales. Writing such fake and untrue reviews is called as review manipulation. After detecting the Spam reviews from the existing Dataset, a new Dataset is created which doesn't contain spam reviews. At last, a new algorithm is proposed that detects spam reviews more precisely and performs opinion mining using Spam Filtered Dataset. This paper indicates that the proposed algorithm is scalable and performs well irrespective type of dataset used.

[6] In the year 2017, an approach to Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining was made. This paper investigates the customer reviews of a product and compare the products, based on reviews at one place. This paper is different from other one as this technique only mines the information of that product on which the customer have expressed their opinions and checks whether the opinion are positive or negative .In order to achieve the desired result it follows 5 steps. In first step while login the customer will be verified using his/her email id ;in second step mining product features that have been committed by customers, in third step identifying opinion sentence in each review and deciding whether each comment is positive or negative in fourth step it is checked whether the given opinion is fake or not, if it is fake then the email id is blocked however in the last step results are summarized. So the challenge of this paper is to find the spam opinion from the huge amount of unstructured data.

[7] In 2018, a Fake Product Review Monitoring Using Opinion Mining was introduced. In this paper a framework is proposed which is used to detect the fake product reviews or spam reviews by using open mining. In this paper researchers are trying to find out the opinion of a customer through a piece of text. It follows the process wherein the reviews are taken and checked whether the review is related to specific product with the help of Decision tree .Spam dictionary is used to identify the spam words in the reviews. In text mining several algorithm are applied in order to obtain the desired result. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Software which is being provided by the researchers that is introduced in this paper will help the user to pay for the right product. Analysis is made through this software and if fake reviews are found from any IP address consistently then admin users will have the authority to block that IP address. It sends the mail regarding blocked IP address to the user .In this way it will monitor the fake review made on any product and thus will give surety to the user about the products availability on that application and reviews too.

[8] In 2018, A framework for Fake Review Detection: Issues and Challenges was proposed. In this paper an overall issues have been summarized as well as challenges for detection of fake reviews as about fake reviewers have been discussed. The proposed framework discussed in this paper is used to deal with both labelled and unlabelled data. This paper works on the framework which works with 90.19% of accuracy for Supervised learning and 83.70% for Semi supervised learning. This paper outlines the issues and challenges from the literature. All the possible features were explored and discussed in this paper. With the available dataset,this paper carries and compiles the experiment and result. All the possible future directions were discussed in this paper.

[9] In 2019, a study research paper was introduced which was titled as Fake Product Review Monitoring System. This paper came up with the objective to bring the solution to all problems related to detecting the fake review. This paper discuss about the various ways to detect spam reviews in order to the opinion mining .A detail discussion have been made on the existing technique to find out whether the review is spam or not. Different techniques such as IP address tracking and Ontology to detect spam reviews are incorporated to get more accurate results from the opinion mining. Once the spam reviews are detected from the existing Dataset then a new Dataset is created which doesn't contain spam reviews and then the Opinion Mining is performed on the new spam filtered dataset. A new algorithm is proposed that detects spam reviews more precisely and performs opinion mining using spam filtered data.This paper includes techniques such as Ontology, Geo location and IP address tracking ,Spam words Dictionary using Naive Bayes ,Brand only review detection and tracking account used.

[10] In 2019, a paper was introduced which was titled as GSCPM: CPM based group spamming detection in online product reviews. . In this paper, researchers proposed a novel of three-step method to detect group spammers based on Clique Percolation Method (CPM) in a completely unsupervised way, called GSCPM. First, it utilizes clues from behavioral data (timestamp, rating) and relational data (network) to construct a suspicious reviewer graph. Then, it breaks the whole suspicious reviewer graph into k-clique clusters based on CPM, and we consider such k-clique clusters as highly suspicious candidate group spammers. Finally, it ranks candidate groups by group based and individual-based spam indicators.They used three real world review datasets from Yelp.com to verify the performance of the proposed method. Experimental results showed that the proposed method outperforms four compared methods in terms of prediction precision. . However, it is advisable to label a small amount of data to improve the precision of detection. Future direction is directed towards the use of partially supervised techniques.

[11] In 2019, Intelligent Interface for Fake Product Review Monitoring and Removal was introduced. In the proposed work a dataset is developed that contains Urdu and Roman Urdu reviews, since it is difficult to detect fake reviews by yourself so , n- gram approached is used to detect fake reviews for multiple languages. While studying it is observed that text categorization with SVM classifier is best approach for the detection of fake reviews. In this way fake reviews are separated.

[12] In 2020, a paper was published discussing Declarative Programming Approach for Review Detection. This paper presentation approach to detect speak reviews incorporated behaviour of authors of reviews combine with properties derived from the content of the reviews. It aim to design of white box approach with is becoming a major requirement nowadays in the industry this is due to the fact that there are increasing social concerns about decision made based on personal information. This paper proposes a rule based fake review detection system using Answer Set Programming which is a powerful tool to declare malicious behaviour patterns specified via a variety of constraints. In this way it combines information about the number of reviews and the number of dislikes also information about the number of reviews and also the analysis of the points in time reviews have been returned and qualitative properties of the content based on similarity measures   and derived classification of reviews and product .This paper proposes a way to create a recognition system for detecting fake reviews, spams and spammers on the web producing fake reviews.

III. OUR FINDINGS FROM STUDIED LITERATURE

The identified challenges motivate to bring up a solution to all the problems stated in the above problem statement section. Following are the objectives of the studied literature:

- To implement different algorithm to get better Spam Detection i.e.; IP Address, Account used, Negative Word Dictionary using Senti-strength, Ontology.

- Graphical representation of work.

- To deals with 6 different types of Spam Reviews.

- To presents Opinion Mining on Spam Filtered Data.

- To implement Ontology in Spam Detection.

- To present an algorithm that does Opinion Mining with Spam Detection.

- Issues and challenges were outlined from literature.

- A framework has been proposed to deal with the issues.

- Possible features were explored and discussed.

- Possible future directions were discussed.

IV. CONCLUSION AND FUTURE WORK

Finding the opinion spam from huge amount of unstructured data has become an important research problem. Now business organizations, specialists and academics are putting forward their efforts and ideas to find the best system for opinion spam analysis. Deceptive reviews available in the internet that increasingly affects businesses and customers. Hence it is important to detect and eliminate such fake reviews from online websites. This paper reveals several approaches used for review spam detection and performance measures were identified. In this paper, various ways to detect Spam Reviews in order to be more accurate and useful have been studied. A detailed discussion about the existing techniques, to find out the whether the review is spam or not is presented. A direction for future research is to implement the system and check performance by applying proposed approach to various benchmark data sets. Comparing performance of different classification methods to find the best one for our proposed opinion spam classification method could be another future research direction. However, there exist other kinds of review or reviewer related features that are likely to make a contribution to the prediction task. In the future we will do further investigate different kinds of features to make more accurate predictions.

V. DISCUSSION

Existing literature is summarized based on the content of the relevant research papers. The problem of fake review detection is a complex problem. This is due to the fact that fake reviews are related to at least two major sources: first, the content of the review and, second, the behavior of the reviewer. Finally, the findings are identified for the sameresearch papers. Gist of the studied research papers is presented in brief. In particular, this paper provided the direction to solve thefake/spam review detection problem through various approaches. Presented work is very useful to the beginners.

REFERENCES

[1] Guan Wang, Sihong Xie, Bing Liu, Philip S. Yu, "Review Graph based Online Store Review Spammer Detection", 2011 11th IEEE International Conference on Data Mining, pp. 1242-1247, 2011.

[2] Huayi Li∗, Zhiyuan Chen∗, Bing Liu∗, Xiaokai Wei∗ and Jidong Shao†, "Spotting Fake Reviews via Collective Positive-Unlabeled Learning", 2014 IEEE International Conference on Data Mining, pp. 899-904, 2014.

[3] Tanya Gera, Deepak Thakur, Jaiteg Singh, "BILD Testing for Spotting Out Suspicious Reviews, Suspicious Reviewers and Group Spammers", 2015 Fifth International Conference on Communication Systems and Network Technologies, pp. 976-981, 2015.

[4] Zeinab Sedighi, Hossein Ebrahimpour-Komleh, Ayoub Bagheri, "RLOSD: Representation Learning based Opinion Spam Detection", 2017 3rd Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pp. 74-80, 2017.

[5] Neelam Duhan, Divya, Mamta Mittal, "Opinion Mining using Ontological Spam Detection", 2017 International Conference on Infocom Technologies and Unmanned Systems (ICTUS'2017), pp. 557-562, 2017.

[6] Manleen Kaur Kohli, Shaheen Jamil Khan, Tanvi Mirashi, Suraj Gupta, "Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 1, pp. 106-109, January 2017.

[7] Anusha Sinha, Nishant Arora, Shipra Singh, Mohita Cheema, Akthar Nazir, "Fake Product Review Monitoring Using Opinion Mining", International Journal of Pure and Applied Mathematics, Volume 119, No. 12, pp. 13203-13209, 2018.

[8] Jitendra Kumar Rout, Amiya Kumar Dash, Niranjan Kumar Ray, "A Framework for Fake Review Detection: Issues and Challenges", 2018 International Conference On Information Technology (ICIT), pp. 7-10, 2018.

[9] Piyush Jain, Karan Chheda, Mihir Jain, Prachiti Lade, "Fake Product Review Monitoring System", International Journal of Trend in Scientific Research and Development (IJTSRD), Volume 3, Issue 3, pp. 105-107, Mar-Apr 2019.

[10] Guangxia Xu, Mengxiao Hu, Chuang Ma, Mahmoud Daneshmand, "GSCPM: CPM-based group spamming detection in online product reviews", 2019.

[11] Ata-Ur-Rehman, Nazir M. Danish, Sarfraz M. Tanzeel, Nasir Usama, Aslam Muhammad, Martinez-Enriquez A. M., Adrees Muhammad, "Intelligent Interface for Fake Product Review Monitoring and Removal", 2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE) Mexico City, Mexico, September 11-13, 2019.

[12] Nour Jnoub, Wolfgang Klas, "Declarative Programming Approach for Fake Review Detection", 2020.
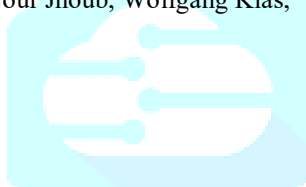
BIOGRAPHIES

Miss Rutuja B. Ardak is pursuing M.E. in Computer Science & Engineering from Sipna College of Engineering and Technology, Amravati. She has received the B.E degree from Sant Gadge Baba Amravati University, India, in 2020. She has granted work in her U.G. project B.E level entitled "Detection of Malicious URL Using Machine Learning". Her area of interests is Data Mining and Embedded Systems.

Prof. Girish S. Thakare has received his M.E. degree in Information Technology and done Ph.D from Sant Gadge Baba Amravati University(MS), India. He is working as anAssociate Professor in Sipna College of Engineering and Technology, Amravati. He has 16 years of teachingexperience. His Research area was Real Time System Scheduling with Energy Harvesting. His areas of interest are Real Time System, and Object Oriented Programming.