

Survey on Interface For Fake Product Review Detection, Analysis And Removal

Vidhya V¹, Albin K Rajan², Abhaynath M R³, Amritha Babu⁴, Mariya Varghese⁵

¹Assistant Professor, Dept of Computer Science

^{2,3,4,5}Dept of Computer Science

^{1,2,3,4,5} MG UNIVERSITY, Depaul Institute of Science and Technology, Angamaly, India

Abstract- *The shift towards shopping online is increasing worldwide day by day. People prefer online shopping over traditional in store shopping for the many advantages which comes in ancillary. The main reason for the boost in online shopping is the convenience it provides to the customer. The customer can switch stores and products at the click of the button rather than traveling to a store. But, there is a catch. The customer could face difficulties when it comes to the quality standards of the product. Customer generated ratings and reviews helps the customer to understand the quality of the product before buying the product. This makes reviews and rating very important aspects that affects purchasing decisions. Therefore a negative review over a good product can affect its sales. A fake negative review can lead to have far reaching consequences. To solve this problem we propose Fake Review Analysis and Detection System (FRADS).FRADS is an Intelligent Interface that uses the Uniform Resource Locator(URL) related to products of the online shopping sites which it monitors and analyzes the reviews provided, and provides the customer with the original rating. The uniqueness of the technology is it can analyze ratings and reviews and also ranks the best reviews in a top down fashion. The proposed system achieved the accuracy of 87% in detecting fake reviews.*

Keywords- Fake Reviews Detection, Text Classification, Natural Language Processing, Machine Learning,Term Frequency and Inverse Document Frequency, Sentimental analysis, Web Crawler, Review Ranking.

I. INTRODUCTION

Earlier people solely depended upon conventional shopping method in which the customer has to visit a particular store or mall for the purchase of a product. Here the customer has no other means but had to rely on the feedback provided by the seller to get any information regarding the quality of a particular product. The seller's inner motive to sell the product and make profit might affect the genuinity of the feedback which he/she provides to the customer And if the customer didn't pay the attention needed while buying the product then it may be proved to be a waste.[1]. On the other

hand, nowadays method of Shopping has changed. Now the customer can also go shopping online. Shopping online provide the customer with additional benefits as he/she is able to buy the same product from online stores run by different brands. But, online shopping doesn't provide customer with the facility to physically examine the product but only provides images and details with respect to the product.

The customer has to make decisions by reading the reviews and buy the product. This shows the importance of product reviews and ratings. The reviews and ratings seen may be fake too. A customer can buy an original and reliable product only if he/she gets the original feedback relating to the particular product. Research in this field shows that U.S. shoppers spend \$6 billion in Black Friday sale 2018 [2]. Americans spend 36% of the shopping budget online. In 2017, E-commerce stores earned \$2.3trillion in sales and is expected to reach \$4.5 trillion by 2021.Today, the number of e-commerce stores around the world is almost about 12-24 million. And studies have also shown that 61% of Amazon reviews that belongs to Electronics Category are fake [3]. These show us the growth of online shopping and the importance of this research. There are some websites which can detect fake reviews like Fake spot. Fakespot is an online Website that helps to detect fake reviews using suspicious patterns and reviewers activity is an example for such websites. For the customer to read all the reviews regarding the intended product which he/she plans to buys is a very time consuming process. And there is also the possibility that some of the reviews might be fake.However, if one can provide system which is enough intelligent to understand and differentiate original reviews from fake reviews and ratings for a product might be a source of reliability for the online stores and customers.

In the proposed technique, the reviews relating to a particular product for which the URL is provided are extracted. Then,the system find out the reviews which are fake and then by analyzing all these reviews understands which ones are the original reviews. Previously researches used to detect fake reviews using different approaches which includes , machine-learning approach, opinion mining and sentiment

analysis, identification address etc. Therefore, we have proposed Fake Review Analysis and Detection System (FRADS). Which helps customers to get original items from online stores at the very minimum time along with the original reviews relating to the product. This system provides the customer with reviews written by genuine reviewers. FRADS also helps to filter out fake reviews from large merchant sites where each product gets hundreds of reviews which will help you to decide whether to buy the product or not.

II. LITERATURE REVIEW

There is very rapid increase in the number of fake reviews being posted on online web stores as reviews are having positive and negative impact on the sales of the product.[3] Finding groups of fake reviewers helps in easy spotting of fake reviews and individual fake reviewers who post fake reviews in e-commerce websites. Posting of fake reviews helps in promoting and demoting the product's sales. One method for this is the usage of "Frequent Itemset Mining (FIM)"[5]. FIM helps in finding the relationship among the fake reviewers groups using behavioral and relational models. The groups thus found are called "spammer groups". With the help of expert human judges labeled datasets are formed. In this system the novel relation based model called GSRank to find the relationship between the spammer groups. The technique uses set of item I which contains set of reviewer Id and also each transaction is marked with a set of Id's the ones who post reviews for a particular product. The system then checks using the FIM method to find the group which posts multiple reviews at a time.

The fake review detecting system of Yelp[6] takes in the filtered reviews and do filtering and analyze these reviews.

It uses 2 main approaches for filtering the reviews: Supervised and Unsupervised. When it comes to features, it has behavioral features and linguistic features. It used the work of Ott that uses Amazon Mechanical Turks(AMT) as the starting point. AMT finds the Turkers who post fake reviews on hotels. The system does Fake Review Detection through Sentimental Analysis using Supervised Learning Techniques is proposed with the aim to classify the reviews into positive and negative polarity by using the machine learning algorithm. The proposed system here uses sentimental analysis and text classification methods.

Fake review detection used in Yelp [7] helps to filter faker review from original reviews Extracting predictive features from the posted reviews is the most difficult part of the project and they extract two types of features: review-

centric feature and reviewer-centric features. Here they first count the percentages of each unigram and bigram tokens for fake and non-fake reviews. Then they take the top 100 unigrams and bigrams which have the most different percentages in fake and non-fake reviews. The second approach here leads to the better performance because it processed all the unigrams and bigrams. They test multiple algorithms of machine learning using the Neural Networks and they achieve the highest accuracy of about 81.92%. The System is good in finding the fake reviews but it still needs an improvement in accurately filtering the reviews. The proposed system takes review texts and finds output whether it is reliable or not and also accurately filters the reviews using the combination of various techniques.

Temporal pattern discovery [8] is used in detection of spam reviews where the arrival pattern is checked. For normal reviews the arrival pattern will be stable and correlated to rating pattern. But for spam attacks the pattern is positively or negatively correlated to the pattern. Data set used for this research was a review website on October 6, 2010. Rating, postdate of the review and whether it is Spammer Review (SR) or not was checked for each review collected. Then in the evaluation process they selected 53 stores each which had more than 1000 reviews. The help of human evaluators were used to make sure to conclude whether it had a SR spam attack or not. The system had a precision of 61.11%. The proposed system is having an advantage in finding the fake reviews as it having a better precision percentage.

Ranking technique [9] is used in product model with the use of weights applied to product reviews for the calculation of product review score. Here in this system the reviews which are not related to the quality of the product will be removed. The preprocessing is done through Support Vector Machine (SVM). According to this system, firstly the removal of comments which are not related to the quality of the product is done. At second stage, weights are provided according to votes received to each review. Finally the overall ranking of the product is done. Factors like relevance of the review with product quality, review content and posting dates are taken into consideration. A 10-fold cross validation is done on the training set. Two measures are used to quantify the effectiveness of the ranking models are used. Namely, correlation between the ranking method and Mean Average Precision (MAP). The system uses only two properties for fake review detection. But, as per our future work uses more number of properties in effective finding of fake reviews accurately.

The Fake review monitoring [10] system focuses on the detection of spam reviews using sentimental analysis. Here it removes vulgar and curse words.

Herein the existing system, scrapping of data is done using web crawler. In the preprocessing stage, the conversion of data into the format required is done. The fake reviews are then removed from the mixture of original and spam reviews. Fake Review Detector is used for this purpose. However, the dataset which is used for training the model is very small which cannot find the suspicious patterns more accurately. The proposed data set is comparatively very large hence every kind of patterns will be very well noticed.

III. CONCLUSION

In the proposed work, dataset is developed that contains reviews from the online store specially developed for this system. Detection of fake reviews for every individual customer by oneself is a difficult task which is very time consuming too. Text categorization using SVM classifier is one of the best approaches for the detection of fake reviews. With the growth in online shopping. The number of applications and websites by e-commerce enterprise is also rapidly increasing. In all these sites there are millions of reviews available. And there are also so many organizations focused only providing fake reviews in order to increase or decrease ratings of a particular product. Therefore, the system proposed that detects the fake reviews, classify the reviews which are genuine. It also ranks the positive genuine reviews. It helps the user to get the products from the e-commerce website with at most satisfaction on their mind and pay for the good quality product. As, there are a lot of e-commerce stores which is very relevant in current scenario. The proposed system is having a very great importance in helping the customer to buy genuine and quality products. And also the online shopping websites to solve the problem of fake review detection, analysis and removal.

REFERENCES

- [1] Sinha, N. Arora, S. Singh, M. Cheema, and A. Nazir, "Fake Product Review Monitoring Using Opinion Mining," vol. 119, no. 12, pp. 13203–13209, 2018
- [2] Torbet, Georgina. "U.S. Customers Spent over \$6 Billion on Black Friday Purchases." Digital Trends, Digital Trends, 25 Nov. 2018, www.digitaltrends.com/web/shopping-totals-blackfriday/.
- [3] Sterling, Greg. "Study Finds 61 Percent of Electronics Reviews on Amazon Are 'Fake'." Marketing Land, 19 Dec. 2018, marketingland.com/study-finds-61-percent-of-electronics-reviews-on-amazon-are-fake-254055.
- [4] K. Khan, W. Khan, A. Rehman, A. Khan, Asfandyar. Khan, A. Ullah Khan, B. Saqia, "Urdu Sentiment Analysis," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 9, 2018.
- [5] A. Mukherjee, B. Liu, and N. Glance, "Spotting Fake Reviewer Groups in Consumer Reviews," 2012.
- [6] Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What Yelp Fake Review Filter Might Be Doing?," Aaai, pp. 409–418, 2013.
- [7] Z. Wang, Y. Zhang, and T. Qian, "Fake Review Detection on Yelp Dataset and features," pp. 1–6.
- [8] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," p. 823, 2012.
- [9] Paper, "Mining millions of reviews : A technique to rank products based on importance of reviews Mining Millions of Reviews : A Technique to Rank Products Based on Importance of Reviews," no. November, 2015.
- [10] V. K. Madhura N Hegde, Sanjeetha K Shetty, Sheikh Mohammed Anas, "Fake product review monitoring," Int. Res. J. Eng. Technol., vol. 05, no. 06, p. 4, 2018