



FAKE ONLINE AUDITS DETECTION USING MACHINE LEARNING

Suraj B. Karale¹, Laxman M. Bharate², Snehalata K. Funde³
^{1,2,3}Computer Engineering, TSSM's BSCOER, Narhe, Pune, India

Abstract

Online reviews play a very important role in today's e-commerce for decision-making. Large part of the population i.e. customers read reviews of products or stores before making the decision of what or from where to buy and whether to buy or not. As writing fake/fraudulent reviews comes with monetary gain, there has been a huge increase in deceptive opinion spam on online review websites. Basically fake review or fraudulent review or opinion spam is an untruthful review. Positive reviews of a target object may attract more customers and increase sales; negative review of a target object may lead to lesser demand and decrease in sales. These fake/fraudulent reviews are deliberately written to trick potential customers in order to promote/hype them or defame their reputations. For this purpose, we propose TRUE REPUTATION, an algorithm that iteratively adjusts a reputation based on the confidence of customer ratings.

Index Terms: decision-making, opinion spam, reputation, robustness, trust, unfair ratings

1 INTRODUCTION

Now a day's, online purchasing is goes on evaluating, product rating is one of the key factors which is affecting on products fame and popularity. While using online shopping channels, consumers share their purchasing experiences regarding both goods and services with other potential buyers via evaluation. The most common way for consumers to express their level of satisfaction with their purchases is through online ratings. The trustworthiness of a

reputation can be achieved when a large number of buyers take part in ratings with honesty. If some users intentionally give unfair ratings to a product, especially when few users have participated, the reputation of the product could easily be manipulated. Ratings can't exactly express user's opinion about product as compare to reviews, so reviews are playing more important role in expressing user's opinion about product.

Online review has become a common practice for people to read online reviews for different purposes. For example, if user wants to buy a product, one typically goes to a review site (e.g., amazon.com) to read some reviews of the product. If most reviews are positive, user is likely to buy the product. If most reviews are negative, user will almost certainly not to buy it. Positive reviews can result in significant financial gains and/or fames for businesses, organizations and individuals. This unfortunately, gives strong incentives for review spamming. Review spamming can be called as false reputation.

Where,

False Reputation: It refers to "illegal" activities (e.g., writing fake reviews, also called shilling) that try to mislead readers or automated opinion mining and sentiment analysis systems by giving undeserving positive opinions to some target entities in order to promote the entities and/or by giving false negative opinions to some other entities in order to damage their reputations. For reviews to reflect genuine user experiences and reviews, such spam reviews should be detected. The prior project work is focusing on detecting fake reviews and indirectly fake reviewers/attackers.

II LITERATURE SURVEY

Hyun-Kyo Oh, Sang-Wook Kim[1] defines the false reputation problem in online rating systems and categorizes various real-life situations in which a false reputation may occur. The understanding of why and when a false reputation occurs helps us establish experimental situations. In order to solve the false reputation problem, we proposed a general framework that quantifies the confidence of a rating based on activity, objectivity and consistency.

M. Eirinaki, M. D. Louta, and I. Varlamis[2] defines the factors affecting online shopping intentions and behavior. Paper also focuses on the affect of rating and reviews on online shopping. This paper concludes user's reviews are more affects on online shopping intentions.

Chien-Liang Liu, Wen-Hoar saio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou[3] describes a movie-rating and review-summarization system. Movie reviews are converted into ratings by feature summarization. LSA algorithm is used for feature extraction, natural language processing (NLP), text analysis, text mining which lastly results in ratings.

Suraj B. Karale, G. A. Patil[4] describes another technique for web extraction in view of the qualities of structure of Web page. The tree is represented in the form of DOM (Document Object Model). While breaking down the page, DOM tree is constructed by parsing the webpage. For selecting text content from newspaper four patterns have been implemented. By using parsing technique selected news is converted into DOM tree.

III SYSTEM ARCHITECTURE

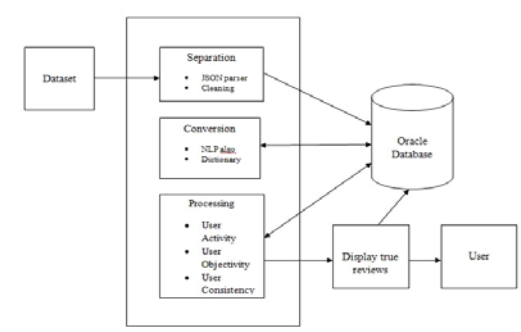


Fig.1: System architecture of online reviews detection.

Fig.1 shows the system architecture of online reviews detection, the input is dataset of reviews which passed to separation module. Separation module filters and cleans the dataset. Filtered and cleaned data get stored in database. The second module conversion will fetch the data from database and required data will be converted to numeric format by using NLP algo, further comparison will be carried out by using word net dictionary. Converted data will be stored in database. Lastly by using three parameters user activity, objectivity, consistency, true or false review will be detected and will be stored into database also result will be displayed.

Module 1: Separation

The separation module works as per following steps:

1. Collection of data from dataset:
 - The data is collected from dataset. The dataset consists of summary of reviews and the rating of the product.
2. Parsing using JSON parser:
 - Reviews are parsed using JSON parser to achieve the results like separation of Reviewer id, review, ratings.
 - This data is either in csv format or JSON format.
3. Cleaning using JSON parser:
 - The JSON parser separates and cleans of fields of reviews. The unwanted data is cleared in a cleaning process using JSON parser, where unwanted words are removed.

Module 2: Conversion

The reviews will be in simple text format, which will be read and converted to discrete ratings using the algorithm, sentiment classification and feature based summarization. We propose Naive Bayes classifier method for sentiment classification of reviews, where reviews will be classified into as per meaning of review words. There view words will be identified using part-of-speech (POS) tagging method and the rating will be denoted as per comparison of words with word net dictionary, where word net dictionary will be a set of ratings as per meaning of keywords. An identification of review words will be used for feature product based summarization. Product features are going to extract using a novel approach called Latent

Semantic Analysis (LSA) and frequency based approach for making group of words.

Module 3: Processing

true reputations, which will be used for desirable product recommendation. The main logic of proposed system will be, if any user satisfies following three parameters then user will be considered as an attacker.

1 User activity

The user activities will quantified as the total number of his ratings for any products. This parameter specifies the how much user is active. User will be called as active user if he will be giving reviews for all products.

2 User objectivity

A rating will be considered more objective if it will be closer to the public’s evaluation (i.e., a reputation) so users objectivity. The objectivity of a rating will be defined as the deviation of the rating from the general reputation of the item. The more similar are the rating and the reputation, the higher will be the objectivity of a rating.

3 User consistency

How consistent the user will be in rating products; in other words, how consistently he will be giving his objectivities of ratings.

IV. Implementation Details

This chapter contains snapshots of proposed system.

4.1 Results

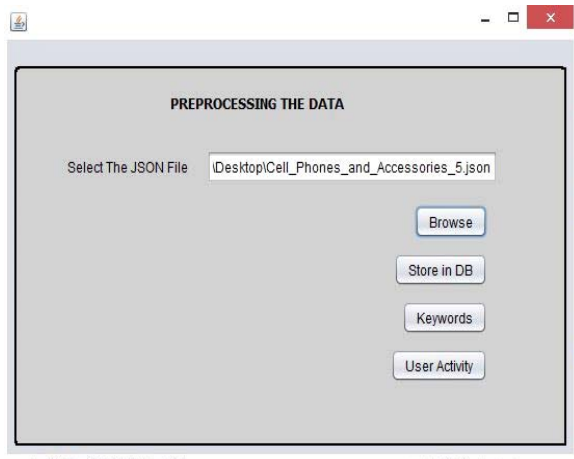


Fig.2: Preprocessing

Fig.2 shows preprocessing of data which contains the file chooser option to browsing the file, that file is further separated and stored in database. The separated reviews are parsed and cleaned to remove stop words and keywords are

The processing will be consisting of three modules, which follows the process of analysis. Analysis will produce a result as set of

separated. This user activity process is carried out to activeness of user.

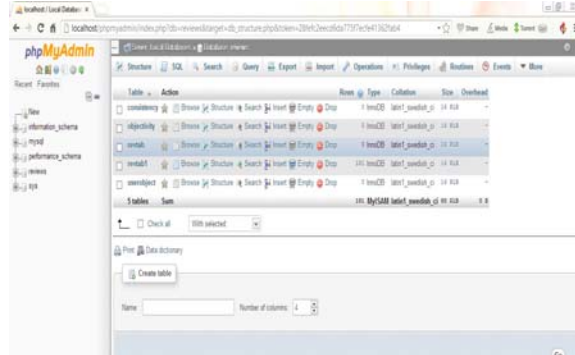


Fig.3: Database

Fig.3 shows database and the content in database. A review is a database created to store review of users. It contains tables such as revtab1, userobject, objectivity, consistency. These tables are used for storing parsed, separated data and results.

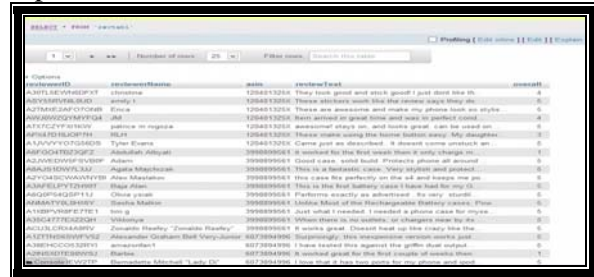


Fig. 4: Separation and parsing process

Fig.4 shows separation and parsing process done by JSON parser. Dataset is separated and stored in table revtab1 as per columns reviewerID, reviewerName, asin, reviewText and overall. Separation is carried out with creating objects with JSONObject as per column names, further data is stored into column using JSONArray function.

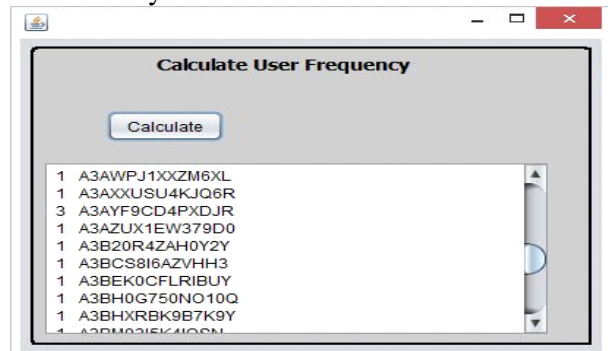


Fig.5: User Activity

A user who posts many reviews should be considered an active user. Online reviews systems often include extremely active users who provide a comparatively large number of reviews. Fig.4 represents how user is active, how many times reviews have been given by him. It is calculated by count(*) and group by function with respect to reviewerID, which is taken as au.



Fig.6: UI User Objectivity and Consistency

Fig.6 is a UI of user objectivity, consistency and true rating. It represents users objectivity, how user is consistent to posting a reviews and true reputation respectively.

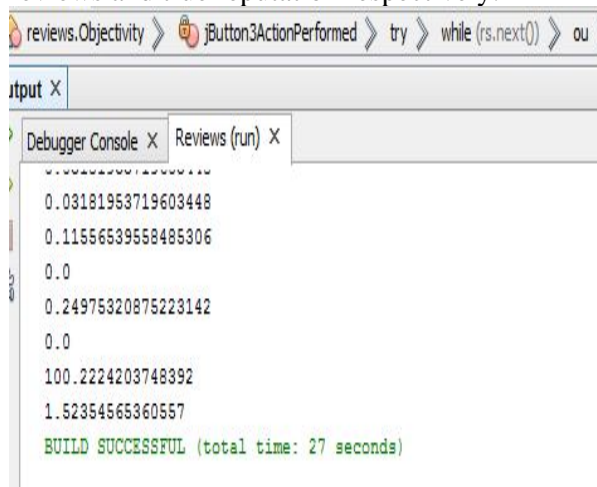


Fig.7: True ratings

Fig.7 shows true reputation of reviews. This true reputation find by using three parameters i.e user activity, user objectivity and consistency. True rating is calculated by $au*ou*cr$ (au=user activity, ou=user objectivity, cr=user consistency).

V. CONCLUSION

5.1 Conclusion

In the present scenario, customers are more dependent on making decisions to buy products either on ecommerce sites. Since these reviews are game changers for success or failure in sales

of a product, reviews are being manipulated for positive or negative opinions. Manipulated reviews can also be referred to as fake/fraudulent reviews or opinion spam or untruthful reviews. In today's digital world deceptive opinion spam has become a threat to both customers and companies. We have implemented a system which can be able to find true reputation, which will help buyers to determine true reputations of a product also actual trustworthiness of a product.

Thus system also represent the various algorithm, tools, language which help us to calculate user activity which shows how much user is active, calculate user objectivity which checks the objectivity and last user consistency which checks the how user is consistent. This system is use for recommendation of product. It is use to find fake user or paid user and also finds the activeness and consistency of user. In future the e-commerce sites can use this system for online and upcoming reviews.

REFERENCES

1. Hyun-Kyo Oh, Sang-Wook Kim, Member, IEEE, Sunju Park, and Ming Zhou, "Can You Trust Online Ratings? A Mutual Reinforcement Model for Trustworthy Online Rating Systems" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, YEAR 2015.
2. M. Eirinaki, M. D. Louta, and I. Varlamis, "A trust-aware system for personalized user recommendations in social networks," IEEE TRANSACTION ON SYSTEM, MAN, CYBERN., SYST., VOL. 44, NO. 4, PP. 409–421, APR. 2014.
3. Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou, "Movie Rating and Review Summarization in Mobile Environment", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 42, NO. 3, MAY 2012.
4. Suraj B. Karale, G. A. Patil "Extracting brief note from Internet Newspaper" IEEE Transaction on Computing for

- Sustainable Global Development, vol. 45
no. 1 March 2016.
5. E. P. Lim, V. A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proc. 19th ACMInt. Conf. Inf. Knowl. Manage. (CIKM), Toronto, ON, Canada, 2010, pp. 939–948.
 6. P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, "Providing justifications in recommender systems," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 6, pp. 1262–1272, Nov. 2008.
 7. Kolli Shivagangadhar, Sagar H, Sohan Sathyan, Vanipriya C.H., "Fraud detection in online reviews using machine learning techniques", *IJCER*, volume, 05 Issue, 05 May-2015.
 8. J.-S. Lee and D. Zhu, *INFORMS J. Comput.*, "Shilling attack detection—A new approach for a trustworthy recommender system," vol. 24, no. 1, pp. 117-131, 2012.