Detection of Fraud Ranking in Mobile Apps
Deepti S. Shrama¹, Ashwini V. Patil², Renuka D. Thakur³, Dikshita R. Chaudhari⁴
Computer Department
GCE, Nagaon, India

Abstract:
Fraud ranking smart phone apps means fraudulent or manual boosting of app ranking to make it more popular. The app developers and the app owner’s user fraud ways to bring up the ranking in mobile store list. This can include dedicated agencies for boosting the ranking or with the help of social network. This in turn make the mobile user who wishes to download app for specific life more difficult as due to wrong ratings the user is misguided and he/she ends up with a app which actually is not very helpful but yet had a good ranking in app store. This makes user waste his time as well as his/her data bandwidth. This also makes wrong impression in users mind regarding that particular apps category which is due to fraud mobile ranking. Considering this a proposed system will process and predict the fraudulent nature of reviews for specific app and present user with truer app ratings and rankings.

Keywords: Fraudulent, Mobile App, Reviews, Smartphone, Text Mining.

I. INTRODUCTION
Increasing use of smart phones has increase the number of mobile applications. There is smart phone based application for e-commerce, food ordering, mobile games, health and safety, entertainment etc. There is large number of applications for each and every category. So with wide range of apps each and every app is given a rank on the mobile app stores. This rating and ranking are obtained by analyzing the reviews of users for that specific app considering this scenario the app developers and app owners are in serious race to increase their apps ranking in app store. To achieve this they often use fraudulent ways such as boosting their ranking by bulk reviews or hiring agencies which manually boost their app ratings. Even social network can be used to accomplish this. This all is done to increase the popularity of the app. Consider a scenario a large group of friends downloads and reviews a specific app in order to increase its popularity. People also achieve by asking their employees to rate and boost the ranking of that particular app. All this is very good from marketing point of view but this does not help the app downloader who sees the ranking of app and then downloads it. He gets wrong information and he is misguided totally and ends up with either downloading a not so satisfying app whereas other apps were actually more helpful. This is also wrong from the point of view that few organizations are big and they can invest more in the marketing of their apps will end up with high ranking for their app in play store. On the contrary a small organization that actually had developed a very good app which is actually more useful for users ends up with low ranking. So the proposed system aims to detect these fraud rankings in mobile apps and present user with truer picture. The system would use text mining algorithms to evaluate the opinion level of a particular review. It would initially detect reviews fraudulent level and according to it would rank the app. It would also implement ways by which a sudden increase in rating and sudden boosting of app ranking can be detected.

II. PROPOSED SYSTEM

SYSTEM ARCHITECTURE

Considering the problems user face due to mobile app fraud reviews the proposed system aims to minimize the fraud review rating. The proposed system is a client server architecture where user will search for review and rating of specific mobile app from his smart phone and server will process and return the appropriate ratings and review of the application requested by user. User application will be an android application. The server will make sure that all the fraudulent reviews are analyzed and user is provided with truer results. All the reviews related to a specific application will be stored in a centralized database along with reviewer information such as names, organization etc. The user can also review and rate any application from his android mobile application. The system will also make sure that only one review arrives from any mobile for that system will track users ID.
The server will perform a systematic check on the reviews to find review fraud it would include the following:

1) Checking the reviewer in the list of developers other similar application. If the reviewer is common in most of the application of the developer whose application is under processing the reviewer reviews will be given a less weightage.

2) Checking for sudden rise in positive reviews for a particular application: Mostly when developers or owners use fraudulent means to boost their app rating it is found that the reviews arrive in a very short frame. Server will also process and find such pattern and these reviews will be considered as fraudulent.

3) Checking for common terms in reviews: Mostly when developers or owners boost their rating it is found that reviews contains common terms as the reviews are more likely from common source. If this is the case these reviews will also be given less weightage.

4) Checking for duplicate reviews: Many times to boost the rating duplicate reviews are been posted. These reviews will be found out and filtered. Finally once it’s made sure that all the reviews under consideration are been given right weightage according to their trueness polarity of the reviews is found out by comparing the reviews with positive and negative words corpus and the results are returned to user.

III METHODOLOGIES USED

1) Text Extraction/Parsing: -Parsing refers to scanning the entire review and retrieving only the text part from the reviews.

2) Tokenization: -Breaking the entire text into word from is known as tokenization. To tokenize a text delimiters are used such as (? ! ; .) etc.

3) Filtering Tokens: - Filtering refers to remove noise from the text. There are many stop words which are not important and increase the processing time. So filtering firstly scans and removs all the stop words from the list. Secondly words in the text have different form so stemming is applied to it so that all the words all brought to their root form.

4) Opinion Analysis Phase: - In opinion analysis the words are compared with the dataset of positive and negative words to find out the polarity of the review.

IV. RELATED WORK

The detection mechanism is broadly divided into three categories the first is web ranking spam in the paper aleandros at al [1] detects spam using heuristics classification algorithm and that works on the given training data and the set of attributes to create a flow chart like structure where each internal node denotes a test and every branch gives outcome of the test the end node represents the classes. The paper has proposed a technique to detect the content based spam .However this technique works efficiently when combined with some other technique rather than in isolation. The effectiveness of the classifier is tested and found to be 82.2% of spam pages, while flagging very few legitimate pages as spam and further research in this area can give better result.

In this paper Nikita Spirin et al [2] this paper gives a review of certain techniques first that analyzes the content like word count or the language model and content duplication. While others uses link-based information such as neighbor graph connectivity, performs link based trust and distrust propagation ,link pruning ,graph based label smoothing and study statistical anomalies.

In this paper Bin Zhou et al[3] address the problem of unsupervised web spam detection by using the concept of spam city to detect the like hood of page been spam. Spamicity does not uses the training data set as in classification which a supervised method it uses a real data set. This paper proposes a online link spam and Sampicity which is more efficient and based on three measures such as utility-based link spamicity (ULSpam), utility-based term spamicity(UTSpam)and characteristics-based term spamicity(CTSpam).it detects the spam pages with the spamicity scores if the score value is low then the pages are normal otherwise spam pages. The characteristics-based term spamicity measures are combine working on the six heuristic values to achieve effectiveness which is not efficient individually.

In this paper Ee-Peng lim et al [4] proposes a method to detect users generating spam reviews by analyzing the characteristics behavior of the reviewer it performs two activities firstly try to target one product and increase its impact and then divert this product from other reviewer .Spammer are scored to represent the degree of spamicity of the reviewer and then the spammer with higher degree are further analyzed. The paper shows
that the detected spammers have more significant impact on rating as compared to unhelpful reviewer. This behavioral analysis can further be enhanced to improve the accuracy of the spam detection.

In this paper Zhang Wu et al[5] proposes a Hybrid Shilling Attack Detector(HySAD) to solve the problem of labeled and unlabelled profiles available in real data set to detect Shilling attackers. HySAD uses a MC-relief to select the efficient detection metrics, and Semi –supervised Naive Bayes to precisely separate Random Filler model attackers and Average-Filler model attackers from normal users. An Experimental result of the large data set of Amazon.cn has been used to detect the efficiency of the HYSAD System and Justifies the system to be applicable in the real world.

Sihong Xie et al [6] in the paper focuses on the singleton reviews which has the power of subverting the products sell in an electronics ecommerce. To solve this problem it uses an observation mechanism to the normal reviewer whose arrival pattern is stable and uncorrelated to their rating pattern which spam attackers are usually bursty and either positively or negatively correlated to the rating. For this a multidimensional time series mechanism is used where a multi-scale anomaly detection algorithm based n curve fitting is used and its efficiency is tested with the help of an experiment support.

Kent Shi et al [7] in its paper address the problem which is observed in GetJar which shows a few well known apps popular among the user but vast majority of the apps are rarely used by most of the users thus the goal of this proposed system is to provide a top-N list of apps to each users based on her recent apps usage by personalized usage analysis and then the quality of the prototype is evaluate on the bases of the accuracy which represent the ability of the recommender to predict the presence of an app on the users device. The Eigenapp model performs well in terms of accuracy and in promotion of the less well known apps in the tail of the data set. The limitations of the system is that it uses app with certain minimum usage, a condition that most apps satisfies while in this there is a possibility that we would remove the some high quality most interesting app which has not been exposed due to the lack of the marketing however in this current model it is using a context based model to extract useful features from the app metadata plain to combine the results of the collaborative and content based approaches.

Bo Yan et al [8] in his paper suggest a way to find the interesting application in Crowded App store using an AppJoy which is an complete automatic system which does not require any user input and has a potential to change with the requirement of the user. It uses a prediction algorithm to provide the efficient result. When tested this android app shows an 4% of battery usage on Off-the shelf devices, perceive response latency was low about 3-5 seconds on Wifi and 3G connections, the recommendations’ predicted usage score achieved more than 80% accuracy for more than 80% of the user. The AppJoy has not integrated the user context which would have further improve the recommendation algorithm of the AppJoy however it supports a large scale analysis of the real world mobile data.

Hengshu Zhu et al[9] in this paper has given classification of the app in into categories like .it utilizes the web knowledge through the web engines and certain context –rich device logs of the mobile apps to retrieve the contextual information for efficiently classifying mobile apps in the real –world context. And then finally combines this information into a maximum entropy model for training a mobile app classifier.

V. CONCLUSION

Now a days, many of mobile app developers uses various fraud techniques to increase popularity of their apps . To avoid false popularity of the mobile application, this paper proposes a methodology based on text mining on user reviews.

VI.REFERENCES

[1] Marc Najork, Dennis Fetterly, “Detecting Spam Web Pages through Content Analysis”Microsoft Research 1065 La Avenida Mountain View, CA 94043, USA

[2] Nikita Spirin, Jiawei Han, “Survey on Web Spam Detection: Principles and Algorithms”Department of Computer Science University of Illinois at Urbana-Champaign Urbana, IL 61801, USA spirin2@illinois.edu


[5] Zhiang Wu, Junjie Wu2*, Jie Cao1, and Dacheng Tao3“HySAD: A Semi-Supervised Hybrid Shilling Attack Detector for Trustworthy Product Recommendation” Centre for Quantum Computation & Intelligent Systems, University of Technology, Sydney zawu@seu.edu.cn, wujj@buaa.edu.cn, caojie690929@163.com, Dacheng.Tao@uts.edu.au

[6] Sihong Xie, Quan Wang, Shuyang Lin, Philip S. Yu, “Review Spam Detection via Temporal Pattern Discovery” Department of Computer Science, University of Illinois at Chicago, Chicago, IL Computer Science Department King Abdulaziz University Jeddah, Saudi Arabia {s xie6, gwang26, slin38, psyu}@uic.edu,2012


[8] Bo Yan and Guanling Chen “AppJoy: Personalized Mobile Application Discovery” Department of Computer Science, University of Massachusetts Lowell 1 University Avenue, Lowell, MA 01854 {hyan, glchen}@cs.uml.edu,2011

[9] Hengshu Zhu1 Huanhuan Cao2 Enhong Chen1 Hui Xiong3 Jilei Tian “Exploiting Enriched Contextual Information for Mobile App Classification” University of Science and Technology of China 2Nokia Research Center 3Rutgers University 1{zh, cheneh}@ustc.edu.cn 2{happia.cao, jilei.tian}@nokia.com 3hxiong@rutgers.edu,2012