7th National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(1st and 2nd March 2019)
International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

Detection of Spam Mobile Application Using Sentiment Analysis

Manoj Rane[1], Sanyukta Patil[2], Ankit Dhavdekar[3], Janhavi Sangoi[4]

Department of Computer Engineering, Mumbai University, INDIA

Abstract*—Mobile application is a very popular and desired concept due to the rapid development in the mobile technology. For the mobile application users there are about more than 3 million applications available for use. However, most of the mobile users firstly prefer the high ranted apps present on the leaderboard chart for accessing it. But the system proposed here cannot guarantee the consistency for the applications that are download, since there is increasing number of mobile app ranking spams or fraud. In mobile app market ranking fraud refers to the duplicitous of actions which have an intention of bumping up the apps in the popularity list on the leader board chart. Many application developers make use of various promotional or advertisement method to be high on the leader board. However, as a boon instead of relying on the real marketing strategies, shady or tricky app developer resort to some fraudulent means to deliberately boost their apps and eventually manipulate the ranking on app store and earn a million bucks through this process. to stop this fraudulent act the current system uses the following steps which includes collecting the historical records for finding leading events in each evidence and second evidence exaction where all the three evidence i.e. rating, ranking and review are exacted and then aggregated through EIRQ algorithm. Then it finds whether the app is false or not. The aim of proposed system is to enhance the prevention of marketing frauds in mobile app using the evidence extraction and aggregation of all three evidence. The three evidence used are ranking based , rating based and review based. In review based the proposed system uses sentiment analysis using HARN algorithm.*

Keywords—Rating, Review, Ranking evidence, Sentiment analysis, HARN algorithm.

I. INTRODUCTION

The mobile apps are growing rapidly and everyday many apps get launched as well as some of them get closed and in these apps some of them are fraud apps which can completely damage mobile phones. as apps are growing daily and many new apps are launched it is very difficult for the user to select the best app. so app store have leader board to showcase some best app according to the category. There are various ways of advertising apps to promote it on leader board[1]. But some there are illegal ways of bumping the app on leader board. So to restrict this kind of fraud some of necessary points are that an app can be rated only one time from user login and implement with the benefit of IP address that limits the number of user logins.

At long last, the proposed work will be analysis with App information which is to be gathered from the App Store for a long-term period called historical records. In the current work, from the gathered historical records, the main occasion and leading session of an application is distinguished. There are two primary advances for mining leading sessions. To start with, it have to find leading occasions from the App's historical positioning records. Second, it need to consolidate nearby leading occasions for building leading sessions[6]. Cautious perception demonstrates that the portable Apps are definitely not continuously at best most position in leader board. In any case, just in a few time span called leading occasion which is diverse leading sessions implies fraud which especially happens in this leading session. At that point from the client judgmental input, three unique sorts of evidence are gathered in particular positioning based evidence, rating based evidence and based evidence. As our undertaking in view of evidences gathered from application information; the one of the generally judgment by individuals is evaluating based evidences which can be utilized to rate the application while downloading it or it can rate it in the wake of seeing its execution.

As examined above there are a few strategies with help of which the rating can get increments by doing fraud. So, the other refereed evidence based technique is review based evidence, finding to make the exact description of app whether it is good or bad app to download. In Review Based Evidences, other than evaluations, a large portion of the App stores additionally enable clients to compose some literary remarks as App surveys[4]. Along these lines, individuals may beyond any doubt think about downloading that particular application by perusing remarks determined in review section area and furthermore give their

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

conclusion about that application. Because of the tremendous number of applications, it is difficult to look fraud for each applications; along these lines, it is vital to have an adaptable method to naturally identify fraud without utilizing any benchmark data.

II. RELATED WORK

According to the studies R. Mariappan, et. al. [1] have proposed a system for mobile fraud application on ranking, rating and reviewing using statistical hypothesis test. First it shows that ranking fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. H. Zhu, et.al. [2] have submitted a paper of system on discovery of ranking fraud for mobile application. In this paper, the dataset is read and then it is preprocess to separate it out into the textual review and the statistical review A. Tale and R. Shahane [3] have proposed a system on ranking fraud for mobile application. In this paper, they have provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile Apps. Specifically, it first proposes to accurately locate the ranking fraud by mining the active periods, namely leading sessions, of mobile Apps. Such leading sessions can be leveraged for detecting the local anomaly instead of global anomaly of App rankings R. Suchdev, et. al. [4] have proposed a system for twitter sentiment analysis using machine learning and knowledge based approach. The major approach used to classify the tweets is the machine learning technique. There are four steps to acquire the results. First, the sanders analytic dataset is acquired H. Kaur and V. Mangar [5] have explained various sentiment analysis approaches. This paper presents a survey on various sentiment analysis approaches. The two major sentiment analysis approaches used are subjective lexicon and machine learning. Subjective lexicon are collection of words where each word as a source initiating the positive, negative, neutral and objective nature of the given text A. Joshi, et. al. [6] have proposed system for automatic sarcasm detection. This paper surveys approaches to automatic sarcasm detection. The system observed three milestones in the history of sarcasm detection research: semi-supervised pattern extraction to identify implicit sentiment, the use of hash tag-based supervision to create large-scale annotated datasets, and the use of context beyond target text. K. Shirai , et. al. [7] have proposed a paper on recognition of sarcasm in tweets .This paper presents new a method for recognition of sarcasm in tweets. The method is based on a variety of approaches, including sentiment analysis, concept level knowledge expansion, coherence of sentences and machine learning classification. N. Hete and D. Sable [8] have proposed a system on mobile application ranking fraud .This paper firstly presents, the mining leading session is utilized to find out the leading events of app for rating reviewing and ranking based evidence then. The KNN calculation is utilized to enhance the effectiveness and precision of the application. A. Prasad, et. al. [9] have proposed a paper on Sentiment analysis on sarcasm detection. This paper proposes a system which detects the sarcastic and non-sarcastic sentence for sentiment analysis. D. Nagarjun, et. al. [10] have proposed a sentiment analysis using harn algorithm .There are few limitations in sentiment analysis various techniques like semantic ambiguity, comparison sentences, Domain dependency problem and specific to English language The basic problem that has been identified is polarity switching of a word based on the domain in.

III. PROPOSED SYSTEM

Specifically, the system first proposes a simple yet effective algorithm to identify the leading sessions of each App based on its historical ranking records. Then, with the analysis of Apps' ranking behaviors, it will find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Hence, the system will characterized some fraud evidences from Apps' historical ranking records, and develop three functions to extract such ranking based fraud evidences. Nonetheless, the ranking based evidences can be affected by App developers' reputation and some legitimate marketing campaigns, such as "limited-time discount". As a result, it is not sufficient to only use ranking based evidences. Therefore, further it propose two types of fraud evidences based on Apps rating and review history, which reflect some anomaly patterns from Apps' historical rating and review records. For the review-based rating the system will perform sentiment analysis using HARN algorithm. In addition, proposed system will develop an unsupervised evidence-aggregation method to integrate these three types of evidences for evaluating the credibility of leading sessions from mobile Apps.

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

IV. WORKING OPERATION

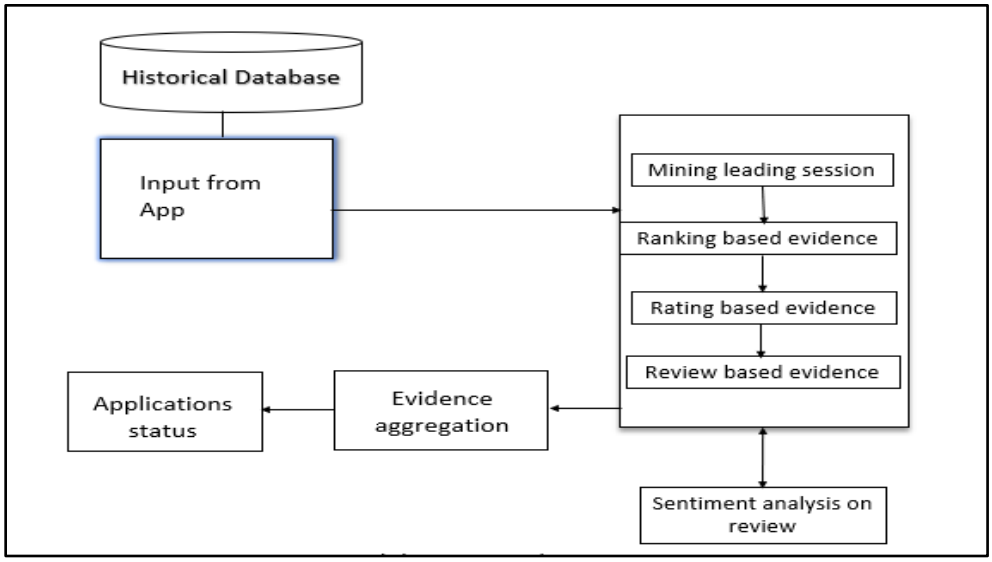

FIG. 1: System Flow Diagram

The above diagram is a system flow diagram for proposed diagram which has following modules: Input from applications, mining leading session of the three evidence (ranking, rating and review), and evidence aggregation algorithm and application status for fraud or not.

4.1 Ranking Based Evidences (Download Based)

A leading session is composed of several leading events, therefore it should first analyze the basic characteristics of leading events of extracting fraud evidences. By analyzing the apps historical ranking records, the system observes that apps ranking behaviors in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase. Specifically, in each leading event, an apps ranking first increases to peak position in leader board, then keeps such peak positions for period(i.e., maintaining phase), and finally decreases till the end of the event(i.e., recession phase).

4.2 Rating Based Evidences

The ranking based evidences are useful for ranking fraud detection. However sometimes it is not sufficient to only use ranking based evidences. Specifically, after an app has been published it can be rated by any user who downloaded it, indeed, user rating is one of the most important features of app advertisement. An app which has higher rating may attract more users to download and can also be ranked higher in the leader board. Thus, rating manipulation is also an important perspective or ranking fraud. Intuitively, if an app has ranking fraud in a leading session, the rating during the time period of session may have anomaly patterns compared with its historical ratings, which can be used for constructing rating based evidences.

4.3 Review Based Evidences

Besides rating, most of app stores also allow users to write some textual comments as app reviews. Such review can reflect the personal perception and usage experience of existing users for particular mobile apps. Indeed, review manipulations is one of the most important perspectives of App ranking fraud. Specifically, before downloading or purchasing new mobile app. Users often firstly read its historical review to ease their decision making and a mobile app contains more positive review that may attract more users to download. Therefore imposters often post fake review in the leading sessions of specific app in order to inflate the app download, and thus propel the app ranking position in the leader board. Although some previous works on review spam

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

detection has been reported in recent years, the problem of detecting local anomaly of review in the leading session and capturing them as evidences for ranking fraud detection are still user explored.

V. RESULT

The application user's device receives data from the server uninterruptedly. Information provided from the application is appropriate according to the users need. The system tells which application on the leaderboard chart is the true rated application. The system perform sentiment analysis on the review-based evidence to give more accurate review.

VI. CONCLUSION

The proposed system develops a ranking fraud detection system for mobile Apps. Specifically, it show ranking fraud happening in leading sessions and provide a method for mining leading sessions for each App from its historical ranking records. Then, it identify ranking, rating and review based evidences for detecting ranking fraud using historical records and statistical hypothesis testIn future plan enhancement for more effective user experience can be done by doing sarcasm detection on review based evidence and app acne be checked through category.

REFERENCES

[1] R. Mariappan, A. Deivanai, Ms. B. Hema, Ms. Thamizharasi ," Investgation Of Ranking Rating And Review Using Statistical Hypotheses Tests", International Journal Of Engineering Sciences & Research Technology ,April 2015.

[2] H. Zhu, H. Xiong, Ge, E. Chen," Discovery of Ranking Fraud for Mobile Apps", IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 1, January 2015.

[3] A. Tule1, R. Shahane ," Discovery of Ranking Fraud for Mobile Applications", International Journal of Engineering Science and Computing, May 2017.

[4] R. Suchdev, P. Kotkar , R. Ravindran , S. Swamy ," Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach", International Journal of Computer Applications Volume 103 – No.4, October 2014.

[5] H. Kaur, V. Mangar, Nidhi, "A Survey Of Sentiment Analysis Techniques", I-SMAC,2017.

[6] A. Joshi, P. Bhattacharya, M. Carman "Automatic Sarcasm Detection", ACM Computing Survey,2017.

[7] P. Tungathamthiti, K. Shirai, M. Mohd "Recognition of sarcasm in tweets based on concept level sentiment analysis and supervised learning approach",28th Pacific Asia Conference Language Information And Computation, October 2016.

[8] N. Hete, D. Sable "A Review Paper On Mobile App Recommendation And Ranking Fraud Detection", International Journal on Research and Innovation Trends in Computing and Communication, October 2015.

[9] A.G .Prasad, S. Sanjan, S.M. Bhat "Sentiment Analysis for sarcasm detection on streaming short text", IEEE Transactions on Knowledge and Data Engineering, Vol. 27, No. 1, January 2017.

[10] D.V. Nagarjana, Devi,T.VRajmikant, P. Rajeshekar "Sentiment analysis is using HARN algorithm" ,IEEE 2016.

[11] P. Shete, S. Dip, V. Kallarkal "Discovery Of Ranking Fraud Detection", International Journal of Recent Trends in Engineering and Research, Volume 03, Issuse 06, June 2017

[12] R. Aggarwal, L. Gupta "A Hybrid Approach For Sentiment Analysis Is Using Classification Algorithm", International Journal of Computer Science and Mobile Computing, Volume 6, Issues 6, June 2017.