Fraud Application Detection Using Data Mining Techniques

Tejaswini Shingare[1], Madhuri Sancheti[2], Swaleha Shaikh[3], Jyoti Ugale[4], Prof.J.N.Kale[5]

BE Students, Computer Engineering, SRES COE, Kopargaon, Maharashtra, India[1]

Assistant professor, Computer Engineering, SRES COE, Kopargaon, Maharashtra, India[2]

ABSTRACT

In today's era, due to rapid development in the mobile technology and mobile devices, the applications i.e. mobile apps are being very interesting and popular concept. As there is large number of mobile Apps, ranking fraud is the challenging factor in front of the mobile App market. Ranking fraud is the term used for referring to fraudulent or suspicious activities having the intention of boosting up the Apps in the popularity list. In fact, App developers are using tricky means frequently for increasing their Apps sales. The main aim is to develop such system that find ranking, rating and review behaviors for investigating review based evidences, rating based evidences and ranking based evidences and then aggregation based on optimization to combine all the evidences for detection of fraud.

Keywords: *Mobile Apps, ranking fraud detection, evidence aggregation, rating and review.*

1 INTRODUCTION

The mobile industry is developing rapidly; therefore the numbers of mobile applications are increasing day by day in the market. As there are many apps available in market users are in fuzzy state while downloading the apps for their use. Different App stores like Google play store and Apple store launched their leader board on daily basis to inspire the users to download most popular applications by observing the ranking of applications. In fact to advertise a particular mobile Apps, leader board of apps is the most important way in the market. An app which is at the top on the leader board leads to large number of downloads and it will gain maximum profit [2]. In order to have their Apps ranked as high as possible, app developers promote their apps using various ways such as advertising, offers etc. Such applications damage to phone and also may cause data thefts. Hence such applications must be identified, so that they will be identifiable for play store users. So we are proposing an android application which will process the information, comments and three reviews of the application with natural language processing to give results. So it will be easier to decide fraud application [5].

2 LITERATURE SURVEY

1. Discovery of Ranking Fraud for Mobile Apps

In this paper, Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen proposed ranking fraud detection system for mobile Apps. Specifically, they first showed that ranking fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. Then, they identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud [1].

2. FairPlay: Fraud and Malware Detection in Google Play

In this paper,Mahmudur Rahman, Mizanur Rahman, Bogdan Carbunar and Duen Horng Chau proposed that FairPlay, a system to detect both fraudulent and malware Google Play apps. They studied on a newly contributed

longitudinal app dataset, in which they had shown a high percentage of malware is involved in search rank fraud; both are accurately identified by FairPlay. In addition, they showed FairPlay's ability to discover hundreds of apps that evade Google Play's detection technology, including a new type of coercive fraud attack [4].

3. MobSafe: Forensic Analysis for Android Applications and Detection of Fraud Apps Using CloudStack and Data Mining

In this paper, Patil Rohini, Kale Pallavi, Jathade Pournima, Kudale Kucheta proposed that most of us use android Mobile these days and also uses the play store capability normally. Play store provide great number of application but unluckily few of those applications are fraud. Such applications dose damage to phone and also may be data thefts. Hence such applications must be marked, so that they will be identifiable for play store users. So we are proposing a web application which will process the information, comments and thee reviews of the application with natural language processing to give results in the form of graph. So it will be easier to decide which application is fraud or not. Multiple applications can be processed at a time with the web application. Also User cannot always get correct or true reviews about the product on internet. So we can check for more than 2 sites, for reviews of same product. Hence we can get higher probability of getting real reviews [5].

4. Survey on Fraud Ranking in Mobile Apps

In this paper, Monali Zende, Aruna Gupta proposed that fraud is happen any time during the whole life cycle of app, so the identification of the exact time of fraud is needed. Due to the huge number of mobile Apps, it is difficult to manually label ranking fraud for each App, so it is important to automatically detect fraud without using any basic information. Mobile Apps are not always ranked high in the leaderboard, but only in some leading events ranking that is fraud usually happens in leading sessions. Therefore, main target is to detect ranking fraud of mobile Apps within leading sessions [2].
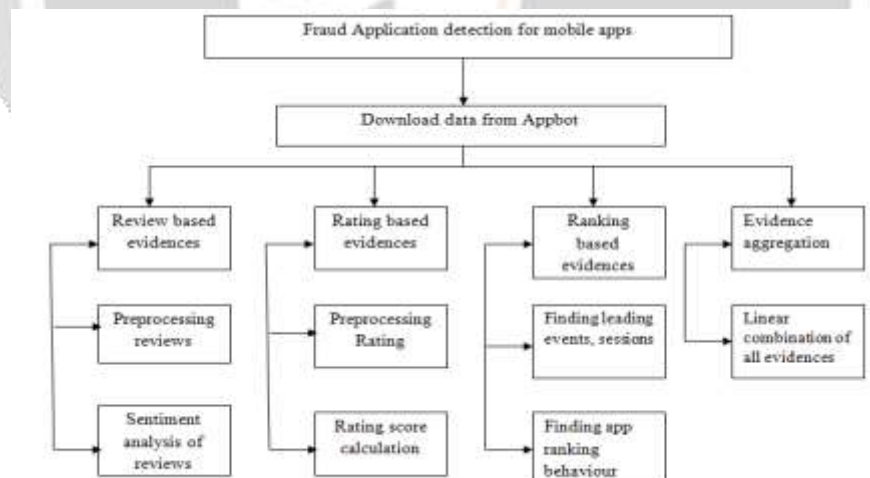
3 SYSTEM ARCHITECTURE

Figure3: Fraud detection system for mobile application

This system consists of four modules described as follows:

1. Rating Based Evidences

2. Review Based Evidences

3. Ranking Based Evidences

4. Evidence Aggregation

3.1 Rating Based Evidences

After downloading an app users generally rate the app. The rating given by the user is one of the most important factors for the popularity of the app. An app having higher rating always attracts more number of users to download it and naturally it can also be ranked higher in the chart rankings. Thus, in ranking fraud of apps, rating based evidences is also an important feature so they are needs to be considered.

3.1.1 Preprocessing of ratings

General ratings are between one to five, in this module it will consider, the rating which are less than or equal to three are considered as negative ratings and rating above three are considered as positive ratings

3.1.2 Rating Score Calculation

Generally, ratings are between one to five, in this module we compute the average rating of particular app and compare it with threshold. The rating which are less than or equal to three are considered as negative ratings and rating above three are considered as positive ratings. Finally, the output is in the form of zeros and ones i.e. negative rating gives zero as an output while positive rating gives one as an output.

3.2 Review Based Evidences

Along with rating users are allowed to write their reviews about the app. Such reviews are showing the personalized experiences of usage for particular mobile Apps. The review given by the user is one of the most important factors for the popularity of the app. As the reviews are given in natural language so preprocessing of reviews and then sentiment analysis on preprocessed reviews is performed. The system will find sentiment of the review which can be positive or negative. Positive review adds plus one to positive score, if negative it will add one to negative score. In this way it will find out score of each of the reviews and determine whether app is fraud or not on the basis of review based evidences. This module contains two subparts given below:

3.2.1 Preprocessing Reviews

This phase consists of following steps:

1. Tokenization: Tokenization is the process of breaking a stream of text into words, phrases, symbols or meaningful elements called as tokens. The list of tokens becomes input for further processing.

 2. Stop word removal: Stop words are commonly used words such as: a, the, and, for, from, is, in and many more.....

 3. Stemming: Stemming algorithm is used to find base word. Porter Stemmer Algorithm is used to find base words.

Porter Stemmer algorithm: Porter Stemmer algorithm is a process for removing suffixes from words in English. Example: A stemming algorithm reduces the words: stems, stemmer, stemming, stemmed as based on "stem".

3.2.2 Sentiment Analysis

After preprocessing of reviews system find out the sentiments of the reviews. It will classify the review as positive or negative. The system will find sentiment of the review which can be positive or negative. Positive review adds plus one to positive score, if negative it will add one to negative score. In this way it will find out score of each of the reviews and determine whether app is fraud or not on the basis of review based evidences.

3.3Ranking Based Evidences

3.3.1 Finding app ranking behavior

In this phase, we detect Apps' ranking behavior, by finding three phases of ranking, namely, rising phase, maintaining phase and recession phase. If the apps ranking reach to peak position in the leaderboard that phase is called as rising phase and maintaining same peak position for specific time period is called as maintaining phase. If the ranking of the app decreases rapidly in the leading event then it is called as recession phase.

Evidence Aggregation

After three types of fraud evidences are extracted, the next work is to combine them for ranking fraud detection. Every evidence is given a Boolean weight as 0 or 1 where 0 indicate no fraud nature and 1 indicate fraud nature.

4 Conclusion

Our system will detect the ranking frauds based on three types of evidences, such as ranking based evidences, rating based evidences and review based evidences. Further, an optimization based aggregation method combines all the three evidences to detect the fraud.

5 References

[1] Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen, *"Discovery of Ranking Fraud for Mobile Apps"*, IEEE Transactions on Knowledge and Data Engineering, Vol.27, No. 1, January 2015

[2] Monali Zende, Aruna Gupta, *"Survey on Fraud Ranking in Mobile Apps"*, IJSR Volume 5 Issue 2, February 2016

[3] Ranjitha R, Mathumitha.K, Meena.S, *"Discovery of Ranking of Fraud for Mobile Apps"*, IJIREM vol.3 Issue-3, May-2016.

[4] Mahmudur Rahman, Mizanur Rahman, Bogdan Carbunar and Duen Horng Chau, *"FairPlay: Fraud and Malware Detection in Google Play"*.

[5]Patil Rohini, Kale Pallavi, Jathade Pournima, Kudale Kucheta, Prof.Pankaj Agarkar,*"MobSafe: Forensic Analysis for Android Applications and detection Of Fraud Apps Using CloudStack and Data Mining",* International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 10, October 2015.

[6]Tejaswini B. Gade,*"A Survey on Ranking Fraud Detection Using Opinion Mining for Mobile Apps"*, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 12, December 2015