RANKING FRAUD DETECTION

[1]Jaydeep Narayan Kale
[1]Assistant Professor,
[1]Computer Engineering Department,
[1]Savitribai Phule Pune University, Pune, India

Abstract— nowadays, due to rapid development in the mobile technology and mobile devices, the applications i.e. mobile apps are becoming very interesting and popular concept in this field. Large number of mobile Apps for different applications are available. But due to huge number of app, ranking fraud is the key challenge in front of the mobile App market. Ranking fraud is the term used for referring to suspicious activities having the intention of boosting up the Apps in the popularity list. Actually App developers are using some fraud ways for frequently increasing their App popularity and sales. The main aim is to develop such system that will detect the fraud nature of App. Rating based evidences and review based evidences can be used for investigating ranking fraud. Once these evidences are identified then they are combined for detection of fraud.

Index Terms— Mobile App, Ranking fraud

I. INTRODUCTION

Today mobile App is a very popular and well known concept due to the rapid advancement in the mobile technology and mobile devices. Due to the large number of mobile Apps, ranking fraud is the key challenge in front of mobile App market. The main aim is to develop such system that will detect fraud and malicious applications on Google play store using three types of evidences i.e. ranking based evidences, rating based evidences and review based evidences [1].

The mobile industry is growing rapidly, subsequently the number of mobile apps coming in the market is also increasing. As there are many apps available in market users are confused while downloading the apps for their use. At the end of April 2013 there are more than 1.6 million Applications at Apples App store and Google Play [1]. Different App stores launched their leaderboard on daily basis to inspire the development of mobile Apps which displays the chart rankings of most popular Apps. In fact for promoting mobile Apps, leaderboard of apps is the most important ways in the market. An app ranking at the top on the leader board ultimately leads to a large number of downloads and million dollars in revenue. User considers that app ranking high in leaderboard is always good but some app developers manipulate these rankings by performing deceptive activities. So the main is to develop such system which will help the user to select the app not only on basis of chart rankings but it will also consider evidences like ratings, reviews and rankings of the app. When maximum of its evidences are good then user can choose that app to use it further. The proposed system will provide prediction about the app being fraud or not using rating, review and ranking based evidences [1].

II. LITERATURE SURVEY

Discovery of Ranking Fraud for Mobile Apps:

In this paper, Hengshu Zhu et al 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. Finally they have proposed optimization based aggregation method to integrate all evidences for fraud detection. [1].

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

In this paper, Patil Rohini et al [2] mentioned 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 cause damage to phone and also result to data thefts. Hence such applications must be marked, so that they will be identifiable for play store users. So they had proposed a web application which will process the information, comments and 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 application 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 they had checked for more than 2 sites, for reviews of same product .

FairPlay: Fraud and Malware Detection in Google Play:

In this paper Mahmadur Rahman et al proposed FairPlay, a system to detect both fraudulent and malware Google Play apps.
They used PCF algorithm to correlate review activities and to combine detected review relations linguistic and behavioural signals gleaned from longitudinal Google Play app data. They used newly contributed longitudinal app dataset. They proved FairPlay's ability to discover hundreds of app that evade Google Play's detection technology, including a new type of coercive fraud attack.

Evolving Recommender System for mobile Apps: A diversity measurement approach:

Xiao Xia et al [4] proposed novel recommendation method utilizing global information about apps. They generated recommendation by both analysing metadata and measuring similarity between apps, leveraging Latent Semantic Index method. They focussed on diversity measurement of mobile app recommendations. New knowledge of user preferences is discovered depending on diversity measurement. They presented preliminary analysis of mobile app recommendation using real time data on large scale.

III. PROPOSED METHOD

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 factor for the popularity of the app [1].

The reviews of the app will be collected and then they can be divided into 3 equal parts. For example if review of certain app is collected for 30 days then they will be divided in to 3 equal parts such that first part will contain reviews of the app for first 10 days, second part will contain reviews for next 10 days and last part will contain reviews for last 10 days. If there is ranking fraud in the selected app then there is more probability of getting higher positive count in one of these parts.

As the reviews are given in natural language so preprocessing of reviews and then sentiment analysis on preprocessed reviews is performed.

Preprocessing Reviews:

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.

Example: This app is Good.

Here,tokens are:

t1 - This

t2 - app

t3 - is

t4 - Good

2. Stop word removal:

Stop words are commonly used words such as a, the, and, for, from, is, in etc.

3. Stemming:

Stemming algorithm is used to find base word.

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.

4. 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.

Rating Based Evidences:

After downloading an app users generally rate the app [2]. The rating given by the user is one of the most important factor 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 need to be considered. 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 a output.

Evidence Aggregation:

After fraud evidences are extracted then next step is to combine them for fraud detection. Every evidence is given a Boolean weight 0 or 1 where 0 indicate no fraud nature and 1 indicate fraud nature.

IV. CONCLUSION

In this paper we discussed various methods for ranking fraud detection. Generally some App developers may use various fraud techniques to increase rank of their app. So to avoid this our paper proposes a method containing modified review and rating based evidences.

REFERENCES

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

[2] PatilRohini, Kale Pallavi, Jathade Pournima, Kudale Kucheta, Prof. Pankaj Agarkar, MobSafe: Forensic Analysis for Android Applications and detection of Fraud Apps using Cloud Stack and Data Mining, International Journal of Advanced Research in Computer Engineering Technology(IJARCET) Volume 4 Issue 10, October 2015.

[3] Mahmadur Rahman, Mizanur Rahman, Bogdan Carbunar and Duen Horng Chau, "Fair-Play: Fraud and Malware Detection in Google Play", arxiv.1703.02002v1 [cs.SI], 6 March 2017.

[4] Xiao Xia, Xiaodong Wang, and Xingming Zhou, "Evolving Recommender System for Mobile Apps: A Diversity Measurement Approach", Smart Computing Review, vol. 3, no. 3, June 2013.