International Journal of Pure and Applied Mathematics
Volume 119 No. 18 2018, 2385-2393
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue
http://www.acadpubl.eu/hub/

A framework for detection of fraudulent apps based on user ranking using incremental learning approach

[1]Mr S.Yuvaraj,[2]Dr Vijay Franklin,[3]Mrs K.V.Kiruthikaa

[1] Assistant Professor, Department of computer science and engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India.

[2]Associate Professor, Department of computer science and engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India.

[3] Assistant Professor (Level II), Department of computer science and engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India.

[1]yuvaraj@bitsathy.ac.in ,[2]vijayfranklinj@bitsathy.ac.in,[3] kiruthikaakv@bitsathy.ac.in

Abstract:

In recent times, the development in the mobile technology and mobile devices, paves way for the mobile application development which become favorite among the developers. In Google Play store, those who have google account access can upload their developed apps in it. As there is large number of mobile apps being uploaded day by day, the importance for ranking, review and rating is increasing and identification of fraud apps is the highly challenging factor in front of the mobile app market. The aim of this paper is to develop a framework that find the behaviors of ranking for various applications with the help of incremental learning approach which includes process such as usage facts, grade facts and evaluation facts and then aggregation of all to determine the ranking fraud detection by classifying the fraud apps and benign apps and recommending to the user. The Evaluation mainly concentrates on ranking given by the user for various applications.

Keywords: Ranking, Review, Rating, Incremental Learning approach.

1. Introduction:

The response given by the consumer plays a major role in assessing the quality of any product in market. Now a day, the online market eases the users work load in many ways and one of the major online resources is usage of applications for various activities and android smart phones is one which used in large numbers around the world. The usage of android mobiles among the users is increasing everyday which opens the gate for application developers in android operating systems. There is need for the development of apps specifically for various applications, that paves way for new developers in the market and the apps creation for android are increased exponentially over the years. The apps count in google play store is also increasing every time because uploading of apps in google play store is cost effective and mostly it requires one-time payment.

The usage of these android apps is become a day to day activity for the smart phone users and it increases the downloads of various apps from the play stores and also it requires frequent updates. This regular process gives way for some fraudulent activities in app development market which raises the questions among the users on genuineity of an app during download. The spammers usually replicate the app which holds major downloads in the market and it will be uploaded in app store which leads to confusion among the user to differentiate between the benign apps and swindle one. The increase in downloads of those fake apps are majorly done by fraudulent reviewers by writing fake reviews to mislead users. To maximize the impact of app download in market which lead to decrease in downloads and usage of original apps by the user. The objective of this framework is to give a fair idea on classification of user's review of an app based on the comments given for a various apps through methods like potter stemmer algorithm and also insight on malware detection.

The paper comprises of following chapters that are arranged as, chapter 2 as literature review, chapter 3 Problem definitions, chapter 4 as proposed system, chapter 5 as Result and discussions and chapter 6 as conclusion.

2. Literature Review:

Rahman, M et al (2017) [1] proposed the idea on fraudulent behaviors in Google's

Android app market fuel search rank abuse. A novel fair play approach is used to find the search rank fraud apps by identifying the trails left by fraudsters. In order to identify the fraud app fair play's PCF algorithm is used to compare the review activities and combines the detected review relation with google play data.

Hao Peng et al (2012) [2] proposed one of android's important mechanisms against malicious apps in which before a user installs an app, warns the user about the permissions the app requires, trusting that the user will make the right decision. This approach is ineffective as it presents the risk information of each app in a "stand-alone" fashion and in a way that requires too much technical knowledge and time.

Junting Ye et al(2015) [3] states that online reviews are an important source for consumers to evaluate products/services on the Internet. Fake reviews are written by the reviewers to mislead users. In order to maximize the impact of the app in play store spam attackes are organized by the spammers.

Leman Akoglu et al (2013) [4] proposed a Fraudeagle framework for finding fraudsters and fake reviews in a online review datasets. Where it concentrates on network effect on the reviewers and products. It uses two complementary steps such as scoring and grouping for fraud detection and visualization respectively. This method is scalable for large data sets.

Michael Grace et al(2012) [5] proposed a mobile phone market reached peak in sales over the years and their popularity leads the malware authors to develop fraud apps .The authors propose a proactive scheme to spot zero day android malware without the use of malware samples and assess the potential security risks.

Hengshu Zhu et al(2015) [6] proposed the holistic view of ranking fraud and fraud detection system for mobile apps by using the leading sessions through mining the active periods of the mobile apps. These leading sessions will be used for the detecting the local anomaly in app ranking instead searching those anomaly globally.

Sudheer Kumar et al (2016) [7] proposed a idea of positioning the misinterpretation and ranking framework for identification of fraud mobile apps. Where the frame work detects web ranking spam, spam detection through online review and app recommendation to the target user.

3. Problem Definition:

The identification of fraudulent apps and malware functionalities is a complex task in the real world. We have many methodologies that are used to prevail over the difficulties faced in this process. But still we have some of the issues in malware identification that are addressed as (a) these algorithms are not suitable for extracting fraud evidences at a particular given time period. In addition, (b) Malware detection is done manually based on manual aggregation. (c) When an App was promoted with the help of ranking manipulation it could be top in leader board and more new users could be purchased that product and (d) it affects the other app's reputations.

With the above issues the insight of primary issues of extracting the fraud evidences by aggregating all the evidences is already proposed to discover the fraud ranking for mobile apps [6].

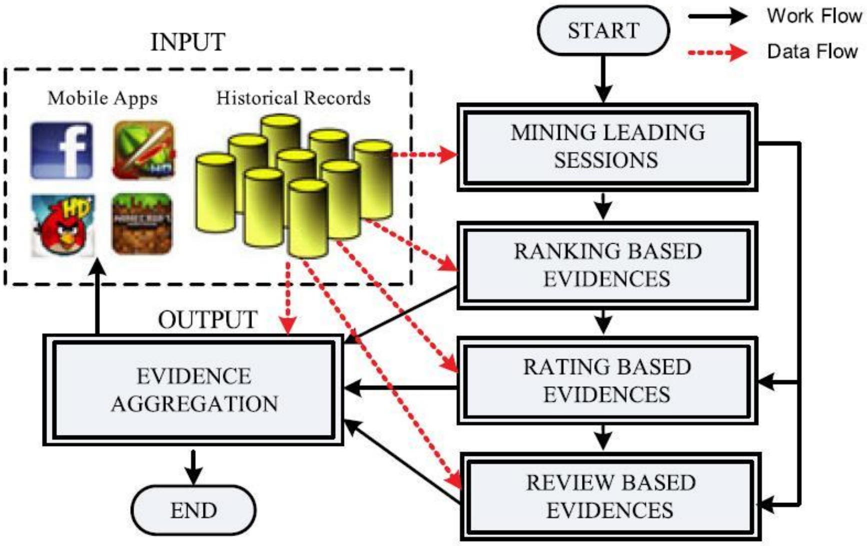

Fig 3.1 System Architecture Diagram for fraud app detection [6]

For finding fraud application in Google Play inputs which are given by the user plays the vital role. Fig 3.1 shows the architecture diagram for finding fraud application which uses the incremental learning approach. In that the apps historical usage information's are collected from user and Aggregate that information into understandable format. Then mining is start with the Leading applications. After finding the leading applications based on historical usage then rating, review and ranking based evidences are calculated.

4. Proposed system:

As the fraud app detection method [7] gives an insight on the evidences based on the review, rating and ranking. The following framework concentrates on the ranking in user perception on the mobile apps to classify the benign app and the fraud app.

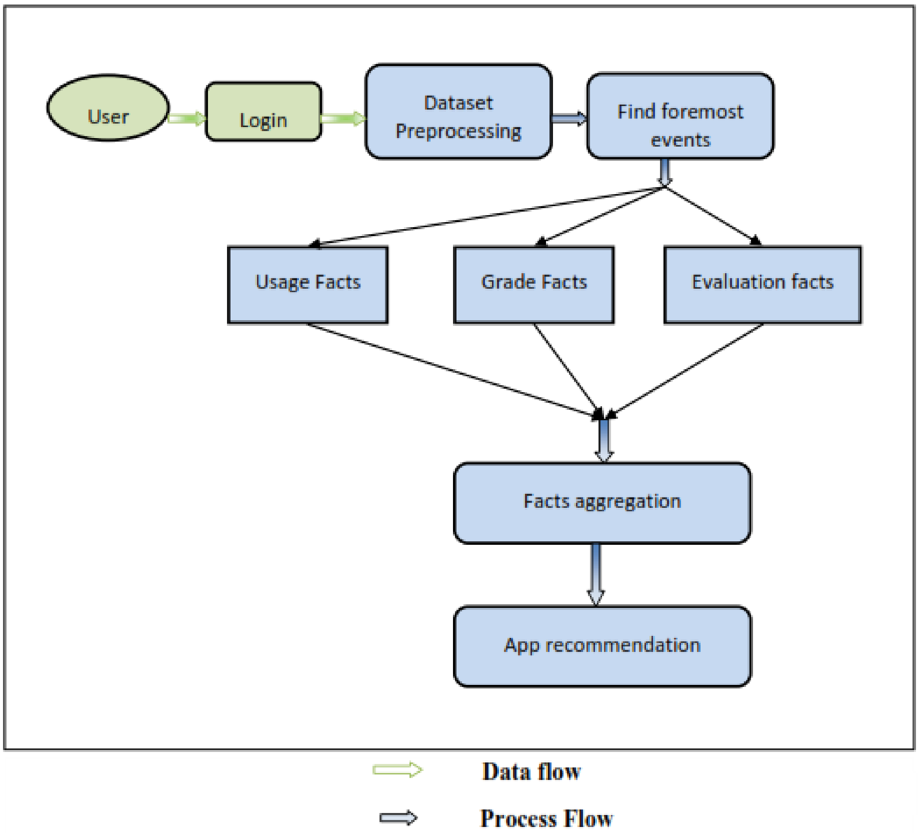

Fig 4.1 Framework for fraud app detection based on user ranking

Procedure:

Step 1: The unordered collection of data is preprocessed to understandable form .which represented in a tabular form.

Step 2: The foremost events are mined to identify the fraud app in which the different patterns of foremost event are compared with normal apps.

Step 3: Foremost event is discovered by historical data of app usage and merging adjacent events for foremost records.

Step 4:a) A Foremost session is composed of several foremost events. The basic characteristics of leading events are analyzed for extracting the fraud evidences. Foremost event consists of three different ranking phases, namely rising phase, maintaining phase and recession phase to satisfy a specific ranking pattern.

b) The ranking is generally used for fraud detection based on the ranking given by the user. But it is not sufficient because 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 of fraud.

c) Apart from rating, the app store allows user to write textual commands as review. Indeed, review manipulation is one of the most important perspectives of app usage facts. Specifically, before downloading or purchasing a new mobile app, users often firstly read its historical reviews to ease their decision making, and a mobile app contains more positive reviews may attract more users to download.

Step 6: After extracting three types of fraud evidences, the next is to combine them for ranking fraud detection. There is not proper method for detecting the ranking of fraud app. Some use supervised learning methods and permutation based models, score based models. In this frame work, unsupervised approach is used based on fraud similarity to combine these evidences.

Step 7: Finally the app is recommended to the user to choose best apps and to avoid fraud apps before to download.

5. Result and Discussions:

To detect the fraud rating, apps dataset is used which consist information around 40000 of android Apps scrapped from Google Play [8]. The fields includes name, date Published, num Downloads Min, file Size, package Name, price, aggregate Rating, software version, rating count, date crawled, url. The dataset is feeded in to the framework and app recommendation is made on the data processed.

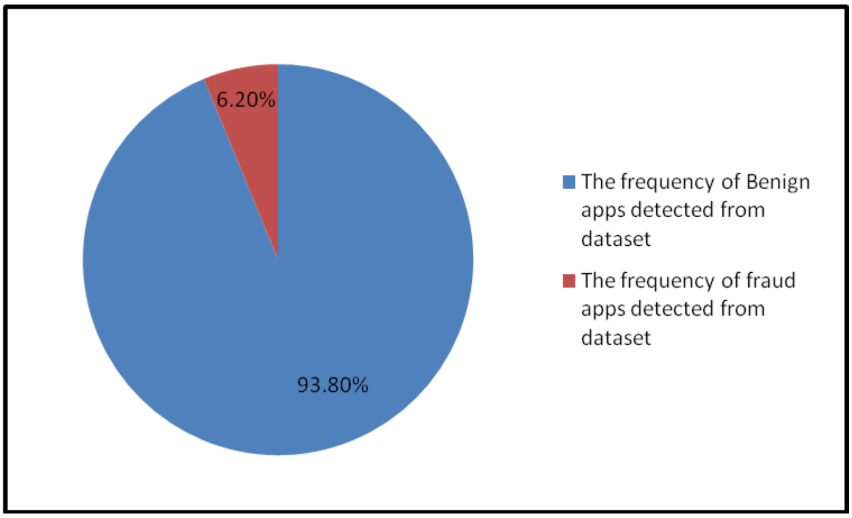

Fig 5.1 Classification of Benign and fraud app based on user rating

By processing the dataset, fig 5.1 gives the classification of Benign and fraud app based on user rating. In which the framework gives the frequency of benign app occurrence is higher than the fraud app. [9]The percentage of occurrence benign app is 93.8% and the percentage of occurrence of fraud app is around 6.2 %.

6. Conclusion and future work:

In this paper, the rating of the user for a mobile app is taken as the primary consideration for deciding the performance of the framework. Apart from rating the review and ranking are also plays the major role in aggregating all the evidence to decide on the nature of the app in the given dataset. In this only app dataset from google play is considered for the performance of the framework. In future there is a scope to feed multiple datasets to ensure the performance of the framework by classifying the Benign and fraud app for different set of mobile applications.

References:

[1] Rahman, M., Rahman, M., Carbunar, B. and Chau, D.H.. Fairplay: Fraud and malware detection in google play. In Proceedings of the 2016 SIAM International Conference on Data Mining ,2017, pg: 99-107.

[2] Hao Peng, Chris Gates, Bhaskar Sarma, Ninghui Li, Yuan Qi, Rahul Potharaju, Cristina Nita-Rotaru, and Ian Molloy., Using Probabilistic Generative Models for Ranking Risks of Android Apps. In Proceedings of ACM CCS ,2012.

[3] Junting Ye and Leman Akoglu. Discovering opinion spammer groups by network footprints. In Machine Learning and Knowledge Discovery in Databases, Springer, 2015 pg:267–282.

[4] Akoglu, L., Chandy, R. and Faloutsos, C., 2013. Opinion Fraud Detection in Online Reviews by Network Effects. ICWSM, 13,2013, pg:2-11.

[5] Michael Grace, Yajin Zhou, Qiang Zhang, Shihong Zou, and Xux-ian Jiang. Riskranker: Scalable and Accurate Zero-day Android Malware Detection. In Proceedings of ACM MobiSys, 2012.

[6] Hengshu Zhu, Hui Xiong , Enhong Chen , Discovery of Ranking Fraud for Mobile Apps, ieee Transactions on Knowledge And Data Engineering, VOL. 27, 2015, pg:74-87.

[7] sudheer kumar G, sreedhar G.S., identifying ranking frauds in mobile apps, international journal of advances in electronics and computer science, issn: 2393-2835, volume-3, issue-10, 2016, pg:16-19.

[8] Balamurugan.E, jagadeesan.A, "Geographic Routing Resilient To Location Errors", International Journal Of Innovations In Scientific And Engineering Research (IJISER) Volume 5 issue-3, 2018, pg:21-26.

[9] Dataset: https://www.kaggle.com/orgesleka/android-apps.