ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Discovery of Ranking Fraud for Mobile Applications

Arati Tule, Prof. Rahul Shahane

Department of Computer Science and Engineering, Wainganga College of Engineering and Management, Nagpur , Maharashtra, India

Department of Computer Science and Engineering, Wainganga College of Engineering and Management, Nagpur , Maharashtra, India

ABSTRACT: As we know that mobile Application market is in rise as mobile users are in large quantity, smart phone users uses those features of mobile Apps as entertaining purpose, knowledge purpose and so on. So the main moto is how we can make sure that the App which we are about to download is as worthy as we think we get after downloading. So here we make it sure by eliminating the fraud Apps and make place the trust worthy one in front of users. This is done by extracting the historical records of those Apps which we are going to verify, we get those Apps information from play store. Then sorting its data as rating based, ranking based and review based evidences and extract the real one rating, ranking and reviews respectively. And then apply algorithm to extract the wholesome data to put further investigate for trust making on that App. Finally, we evaluate the proposed system with real-world App data collected from the iOS App Store for a long-time period. In the experiments, we validate the effectiveness of the proposed system, and show the scalability of the detection algorithm as well as some regularity of ranking fraud activities.

KEYWORDS: Ranking, Review, Rating based evidences; Pattern Analysis, Semantic based analysis, NLP, Weight based Aggregation Method.

I. Introduction

As we know that App market is on high so as developer of App. But there is no trust those days on those developers to surely get what we expect from them. As the developer get reviews about their App and it is seen by every App user; they make sure by viewing those comments. Some App developers get their App ranked high on the basis of its background and get high rankings such as game-loft which get high ranking in short while as developer of that App is already made best of his Apps. But there may be fraud happened as some developers buy these ratings and rankings; those activities are called as ‘bot farms’ or ‘human water armies’.

As we know that the mobile Apps has grown at vast speed in some years; as for march 2017, there are nearby 2.8 million Apps at google play and 2.2 Apps at Apple Apps store. In addition, there are over 400,000 independent app developers all fighting for the attention of the same potential customers. The Apple App Store saw 128,000 new business apps alone in 2014 and the mobile gaming category alone has competition to the tune of almost 300,000 apps. As ranking of App decided on the basis of trustworthiness means more trust worthy the App is the less ranking it get and from those less ranked App users make sure that App is as good to get downloaded. So as per the evidences find from historical records; the Apps which are ranked high for some period but did not manage its position as we found in statistics of that App graph, we easily notice that App is fraud or not. Also there are comments for the similar App from which we found the good or bad comments of those App. But if those comments are achieved from wrong way then; so here we apply some NLP based technique to find the comments liability on semantic level.

II. Related work

In this project, we characterize some fraud evidences from Apps' historical ranking records, and develop three functions to extract such ranking based fraud evidences. We evaluate the proposed system with real-world App data collected from the Apple's App store for a long-time period, i.e., more than two years. 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, we further 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. Thus, we first propose a simple yet effective algorithm to identify the App's rating and review accordingly from scattered data for each App based on its historical ranking records i.e. filtering techniques. Then next propose is to identify the leading events of each App based on data collected as explained above and sorted those events to leading sessions. Specifically, we first propose 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.

Then, with the analysis of Apps' ranking behaviors, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. In addition, we develop a weight based aggregation method to integrate variance of all evidences for evaluating the credibility of leading sessions from mobile Apps. Then we can divide the sessions of each App into parts to easily judge each session. Then according to App's rating, we can calculate its variance for each session and combine those results into one and calculate it with threshold variance to make result corrected solution. Also same thing done for review based sessions using NLP technique on semantic level; so here we first find the App's comments sentiment by analyzing that is it positive negative, more positive, more neg more negative, inverse positive, inverse negative, or neutral accordingly provide 4,2,5,1,2,4,3 respectively.(here inverse positive means negative and inverse negative means positive) .Then according to App's review calculation based on NLP technique described above we can calculate its variance for each session and combine those results into one and calculate it with threshold variance to make result corrected solution. After that calculate the average of variance of rating and review and get result corrected solution as we compare it with threshold variance.

It is worth noting that all the evidences are extracted by modeling Apps' ranking, rating and review behavior. It shows the framework of our ranking fraud detection system for mobile Apps. The proposed framework is scalable and can be extended with other domain generated evidences for ranking fraud detection. Experimental results show the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity of ranking fraud activities.

Here are the implementation of our project on the basis of extracting features such as ranking, rating and review based evidences. And then how we can have caught the fraud based on NLP based technique on semantic based sentimental analysis test here we explained below:

Identifying evidences for ranking fraud detection:

1.Identifying Leading Sessions:Essentially, mining leading sessions has two types of steps concerning with mobile fraud apps. Firstly, from the Apps historical ranking records, revelation of leading events is done and after that merging of adjacent leading events is finished which showed up for building leading sessions.

Finding the Leading events: Given a positioning limit a fundamental event e of App a contains a period range also, relating rankings of App a; Note that positioning edge for example K * is applied which is typically more diminutive than K here in light of the fact that K might be tremendous (e.g., more than 1,000), and the positioning records past K (e.g., 300) are not exceptionally helpful for recognizing the positioning controls. Also, it is finding that a few Applications have a few nearby driving event which are near one another and structure a main session.

Construction of Leading Sessions: Intuitively, basically the main sessions of mobile application mean the period of prevalence, and so these leading sessions will include ranking manipulation only. Consequently, the issue of identifying ranking fraud is to recognize deceptive leading sessions. Alongside the fundamental task is to extract the leading sessions of a mobile App from its historical ranking records.

2.Ranking based evidences: It concludes that leading session contains different leading events. Consequently by examination of essential behavior of leading events for discovering fraud evidences; furthermore for the application historical ranking records, it is been observed that a particular ranking pattern is constantly fulfilled by application ranking behavior in a leading event.

3.Rating based evidences: Previous ranking based evidences are helpful for identification purpose; however it is not sufficient. Determining the issue of "restrict time reduction", recognizable proof of fraud evidences is planned because of application historical rating records. As we realize that rating is been done after downloading it by the client, and in the event that the rating is high in leaderboard significantly that is attracted by most of the mobile app users. Suddenly, the ratings amid the leading session offers ascend to the anomaly pattern which happens amid rating fraud. These historical records can be utilized for creating rating based evidences.

4.Review based evidences: We are acquainted with the review which contains some textual comments as reviews by application client and before downloading or utilizing the application client for the most part want to elude the reviews given by most of the users. Subsequently, in spite of the fact that because of some past works on review spam recognition, there still issue on finding the local anomaly of reviews in leading sessions. So based on applications review behaviors, fraud evidences are used to detect the ranking fraud in Mobile app as we use NLP technique to find it.

5.With and without NLP fraud checking: The field of Natural Language Processing (NLP) aims to convert human language into a formal representation that is easy for computers to manipulate. Current end applications include information extraction, machine translation, summarization, search and human computer interfaces. While complete semantic understanding is still a far distant goal, researchers have taken a divide and conquer approach and identified several sub-tasks useful for application development and analysis. These range from the syntactic, such as part-of-speech tagging, chunking and parsing, to the semantic, such as word sense disambiguation, semantic-role labeling, named entity extraction and anaphora resolution.

III. PROPOSED ALGORITHM

Step 1: Input: App reviews and rating collection: App ID, Text review and rating.
Step 2: Scattered App with its data to get separated using filtering technique.

-A1 T1 R1		FOR APP	TEXTUAL COMMENTS	RATINGS
-A2 T2 R2		A1	T1,T4,T7	R1,R4,R7
-A3 T3 R3	→	A2	T2	R2
-A1 T4 R4		A3	T3,T5	R3,R5
-A3 T5 R5		A4	T6	R6
-A4 T6 R6				
-A1 T7 R7				

Step 3: Select any App accordingly, Review Analysis is done.
Check if the feedbacks have a common trait,
if(F1 = F2 and F2 = F3 and .... Fn-1=Fn)
Then it means the review is genuine
else
if there is a abrupt shift in the pattern, then the feedback might be non-genuine.
Step 4: If selected App is having 35 reviews, then divide those reviews into sessions.
Suppose we divide reviews as 10 reviews per session the here there were 4 sessions in this selected App.
i.e. 10r,10,r,10r,5r in those 4 sessions; r ->review.
Step 5: Per session rating sum:
According to the App's rating here if divided sessions; such as given below:
For session s1= rating is high, s2=low, s3=high, s4=high then overall rating is high but this rating is in star based we consider it as numerical base. Then by taking its variance and compare with predefined threshold we guess that rating based evidences are fake ratings or not.
Step 6: For textual review NLP based searching App detection:
For NLP based technique,
1. Read all feedback information
2. For each feedback, find action words using POS Tagging and Chunking process

3. Evaluate the sentiment from the feedback and mark the feedback as Good or Bad
4. Divide the feedback into sessions
5. For each session find the feedback obtained, to get the list
S1 F1
S2 F2
S3 F3
.
Sn Fn
Where Si is the session, and Fi is the feedback from that session.
so here we first find the App's comments sentiment by analyzing that is it positive negative, more positive, more neg more negative, inverse positive, inverse negative, or neutral accordingly provide 4,2,5,1,2,4,3 respectively.(here inverse positive means negative and inverse negative means positive).
Then according to App's review calculation based on NLP technique described above we can calculate its variance for each session and combine those results into one and calculate it with threshold variance to make result corrected solution so find out fake or genuine App.
Step 7:After that calculate the average of variance of rating and review and get result corrected solution as we compare it with threshold variance and finding of Liability of App.

IV. Simulation Result

Aim: The aim of this paper is providing a holistic view of ranking fraud and propose a ranking fraud detection system for mobile Apps.

Scope:The scope of this paper is investigating three types of evidences, i.e., ranking based evidences, rating based evidences and review based evidences, by modeling Apps' ranking, rating and review behaviors through semantic basis NLP test on review and without NLP on rating and aggregate all those evidences by weight based aggregation method.
Existing System: In the literature, while there are some related works, such as web ranking spam detection, online review spam detection and mobile App recommendation the problem of detecting ranking fraud for mobile Apps is still under explored. To fill this crucial void, in this paper, we propose to develop a ranking fraud detection system for mobile Apps. Along this line, we identify several important challenges. First, ranking fraud does not always happen in the whole life cycle of an App, so we need to detect the time when fraud happens. Such challenge can be regarded as detecting the local anomaly instead of global anomaly of mobile Apps. Second, due to the huge number of mobile Apps, it is difficult to manually label ranking fraud for each App, so it is important to have a scalable way to automatically detect ranking fraud without using any benchmark information. Finally, due to the dynamic nature of chart rankings, it is not easy to identify and confirm the evidences linked to ranking fraud, which motivates us to discover some implicit fraud patterns of mobile Apps as evidences.
Disadvantages:

- The problem of detecting ranking fraud for mobile Apps.
- Huge number of mobile apps, it is difficult to manually label ranking fraud for each app.

Advantages:

- An unique perspective of this approach is that review based evidences can be modeled by semantic based NLP technique through sentimental analysis tests, thus it is easy to be extended from domain knowledge to detect ranking fraud.
- Identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud.

Data Flow of project:
Here is the flow of the project,
1. Gathering data for reviews and ratings from app store, and other sources
2. Pre-processing data to remove any missing entries (using filtering technique)
3. Semantic matching for finding quality of review (NLP on the basis of Positive, Negative or Neutral comments.)
4. Pattern analysis using machine learning

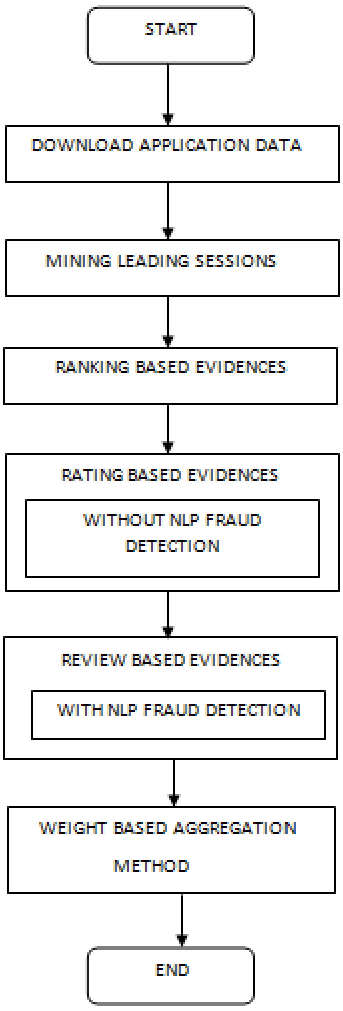
5. Result analysis based on the matching

Gathering data for reviews and ratings from app store, and other sources: To determine the polarity of the sentences, based on aspects, large numbers of reviews are collected from the Web. For judging Apps on without NLP basis ratings are downloaded side by side with reviews also. There are lots of websites on the Internet where the large numbers of customer reviews are available. Amazon website (www.amazon.com) and also play stores like google play are used to collect the reviews.

Pre-processing data to remove any missing entries (using filtering technique): To determine the semantic orientation of the sentences a dictionary based technique of the unsupervised approach is adopted. To determine the opinion words and their synonyms and antonyms WordNet is used as a dictionary; also, it plays a vital role in detecting any missing entries using filtering technique.

Semantic matching for finding quality of review (Positive, Negative or Neutral): A large amount of reviews of users are collected on the Web that needs to be explored, analyze and organized for better decision making. Opinion Mining or Sentiment Analysis is a Natural Language Processing and Information Extraction task that identifies the user's views or opinions explained in the form of positive, negative or neutral comments and quotes underlying the text. Aspect based opinion mining is one of the level of Opinion mining that determines the aspect of the given reviews and classify the review for each feature.

Data Flow Diagram:



DOI: 10.15680/IJIRCCE.2017. 0505054

V. CONCLUSION AND FUTURE WORK

In this paper, we studied ranking fraud detection model for mobile applications. Presently days many of mobile application designers utilizes different frauds systems to build their rank. To avoid this, there are different fraud identification strategies which are concentrated on in this paper. Such procedures are assembled into three classes, for example, web ranking spam recognition, online rating and review spam discovery, mobile application recommendation. Every one of these strategies are viably dealing with ranking fraud detection. Besides, we use weight based aggregation technique to integrate all the evidences variance for calculating exact range that make find out liability of an App by calculating each App's sessions NLP based reviews checking through sentimental analysis test and also without NLP for ratings stars checking for each App's sessions then verify the result corrected solution. A one of a kind point of view of this methodology is that all the evidences can be displayed by statistical hypothesis tests, thus it is easy to be extended with different evidences from domain knowledge to detect ranking fraud. The experimental result shows that propose system save the time as well as memory than the existing system.

REFERENCES

[1] SabbineniPoojitha, Balineni Venkata Sai Mrudula and VemuriSindhura, "A Novel Method To Identify False Apps Through Data Mining", International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 23 Issue 5 –SEPTEMBER 2016.

[2] Ranjitha.R, Mathumitha.K, Meena.S, S.Hariharan, "Discovery of Ranking of Fraud for Mobile Apps", International Journal of Innovative Research in Engineering & Management (IJIREM) ISSN: 23500557, Volume-3, Issue-3, May-2016.

[3] R.Vinodharasi, P.Ramadoss, " Efficient Retrival Of Mobile Apps Using EIRQ", International Journal Of Engineering And Computer Science ISSN: 23197242 Volume 5 Issues 6 June 2016, Page No. 1683016835.

[4] Phopse P.E, Jondhale S.D, "Discovery Of Ranking &Rating Fraud For Mobile Application ", International Journal of Research In Science & Engineering e-ISSN: 2394-8299 Volume 2 Issue 4.

[5] Hengshu Zhu, Hui Xiong, Senior Member, IEEE, Yong Ge, and Enhong Chen, Senior Member, IEEE Discovery of Ranking Fraud for Mobile Apps‖ IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 1, January 2015.

[6] Pranjali Deshmukh, Pankaj Agarkar ―Mobile Application For Malware Detection‖ International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 02 | May-2015

[7] Anuja A. Kadam ,Pushpanjali M. Chouragade ―A Review Paper on: Malicious Application Detection in Android System‖International Journal of Computer Applications (0975 – 8887) National Conference on Recent Trends in Computer Science & Engineering (MEDHA 2015).

[8] Jakub Zilincan ,MichalGregus "Improving Rank of a Website in Search Resuts – a Experimental Approach"2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet

[10] Bing Liu,(2012), "Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers".

[11] B. Pang, L. Lee, and S. Vaithyanathan,(2002),"Thumbs up? Sentiment classification using machine learning techniques" In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86.

BIOGRAPHY

Arati Suhas Tule is in a II year MTech Student in the Computer Science And Engineering Department, College of Wainganga College Of Engineering, Nagpur University. She received Bachelor degree in Engineering in CSE Department 2012 from Raisoni College Of Engineering, Nagpur University.

Prof. Rahul Shahane HOD Information Technology ,WCEM Nagpur

DOI: 10.15680/IJIRCCE.2017. 0505054