International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-03, Issue-04, July 2017

Hybrid Technology to Detect Mobile Application Ranking Fraud

[1]Mr. Satish J. Manje, [2]Prof. Vijay M. Shelake
[1]ME Student, [2]Asst. Professor, [1,2]Comp. Engg. Dept, Mumbai, Maharashtra, India.
[1]satishmanje93@gmail.com, [2]vijaynew12@gmail.com

Abstract— Now a days everyone is using smart phone. Often user requires various applications that can help them get ease their work without adding any difficulty in use. But many try to use this to manipulate their app and making it user friendly without adding much required information. To download an application user often visit Apps store such as Google Play Store, Apples store etc. When a user visit s App store then there are various application which trend to provide same functionality. This list is built on the basis of promotion or advertisement. But many times user do not have knowledge about the app background. So user looks at the list and downloads the applications mostly from front page of App Store. But sometimes it happens that the downloaded application won't work or not useful. That means it is fraud in mobile application list. In this paper it is proposed such mathematical models that can find apps which are fake using some vital information present within an app itself.

Keywords – Mobile Apps, Ranking Fraud, Mining Evidences.

I. INTRODUCTION

The number of mobile app[3] has big at a panoramic rate over the past few years, for instance, as of the top of Gregorian calendar month 2013,there are over one 6 million Apps at Apple's store and Google Play[1]. To stimulate the event of mobile Apps, several App stores launched daily App leaderboards that demonstrate the chart ranking of most well-liked Apps. Indeed, the App leaderboard is one most vital ways in which for promoting mobile Apps. A better rank on the leaderboard typically ends up in an numerous range of downloads and million greenbacks in revenue. Therefore, App developers tend to explore numerous ways in like an advertising campaigns to push their Apps so as to possess their Apps hierarchical as high as attainable in such App leaderboards.

However, as a recent trend, rather than wishing on ancient promoting solutions, shady App developers resort to some dishonest means that to deliberately boost their Apps and eventually manipulate the chart ranking on an App store.

II. AIMS AND OBJECTIVE

AIM

To avoid fraud, this paper creating application during which paper square measure getting to list the application. To list the applying initial this paper square measure getting to realize the active amount of the applying named as leading session. This paper conjointly finance the 3 styles of proofs: Ranking primarily based evidence, Rating proof and Review based proof. Mistreatment these 3 evidences finally this paper shrewd a aggregation. This paper valuate application with planet information collected type play store for lasting amount.

OBJECTIVE

1. To rank fraud for mobile application.
2. To boost the fraud detection potency.
3. This paper ought to first analyze the fundamental characteristics of leading events for extracting fraud evidences.
4. The suspicious leading events might contain terribly short rising and recession phases.
5. This paper ought to analyze net ranking spam detection. Specifically, the online, ranking spam refers to associate degree deliberate actions that rouse select web content an indefensible favorable connection or importance.
6. This paper centered on sleuthing on-line review spam.

III. LITERATURE SURVEY

1] A Survey of Web Spam Detection Techniques Mahdieh Danandeh Oskuie Department of Computer, Shabestar Branch, Islamic Azad University, Shabestar, Iran Seyed Naser Razavi Computer Engineering Department, Faulty of Electrical and Computer Engineering, University of Tabriz, Iran:

Nowadays, with regard to increasing information in web, search engines are considered as a tool to enter the web. Then present a list of results related to user query. A legal way to increase sites rank in the list results of search engines is increasing the quality of sites pages, but this method is time consuming and costly. Another method is use illegal and

unethical methods to increase the rank in search engines. The effort of deceiving search engines is called web spam. Web spam has been considered as one of the common problems in search engines, and it has been proposed when search engines appeared for the first time. The aim of web spam is to change the page rank in query results. In this way, it is placed in a rank higher than normal conditions, and it is preferably placed among 10 top sites of query results in various queries.

2] HySAD: A Semi-Supervised Hybrid Shilling Attack Detector for Trustworthy Product Recommendation:

Shilling attackers apply biased rating profiles to recommender systems for manipulating online product recommendations. Although many studies have been devoted to shilling attack detection, few of them can handle the *hybrid* shilling attacks that usually happen in practice, and the studies for real life applications are rarely seen. Moreover, little attention has yet been paid to modeling both labeled and unlabeled user profiles, although there are often a few labeled but numerous unlabeled users available in practice. This paper presents a Hybrid Shilling Attack Detector, or HySAD for short, to tackle these problems. In particular, HySAD introduces MC-Relief to select effective detection metrics, and Semi-supervised Na¨ıve Bayes (SNBλ) to precisely separate Random-Filler model attackers and Average-Filler model attackers from normal users.

3] Online Review Spam Detection using Language Model and Feature Selection Manali S. Patil PG Student Department Of Information Technology, Pune Institute of Computer Technology, S. No. 27, Dhankawadi, Pune-Satara Road, Pune – 411043.:

In today's age of web 2.0, large numbers of product reviews posted on the Internet. Such reviews are important to customers or users and to companies. Customers use the reviews for deciding quality of product to buy. Companies or vendors use opinions to take a decision to improve their sales according to intelligent things done by other competitors. However, all reviews are given by customers or users are not true reviews. These reviews are given to promote or to demote the product. Some reviews are given on brand of product, and others are related to advertising of another product. There is need to find how many reviews are spam or non spam. In this paper, the system is proposed for detecting untruthful spam reviews using n-gram language model and reviews on brand spam detection using Feature Selection. Given system separately identifies spam and joined the result showing spam and non spam reviews.

4] Rank Discovery From Web Databases Saravanan Thirumuruganathan†,‡, Nan Zhang††, Gautam Das†,‡†University of Texas at Arlington; ††George Washington University:

In this paper, we define a novel problem of rank discovery over hidden web databases. We introduce a taxonomy of ranking functions, and show that different types of ranking functions require fundamentally different approaches for rank discovery. Our technical contributions include principled and efficient randomized algorithms for estimating the rank of a given tuple, as well as

negative results which demonstrate the inefficiency of any deterministic algorithm. We show extensive experimental results over real-world databases, including an online experiment at Amazon.com, which illustrates the effectiveness of our proposed techniques.

5] A Semantic Association Page Rank Algorithm for Web Search Engines Manuel Rojas Oklahoma State University, CS Department mrojas@ okstate.edu:

In this study, I propose a relation-based page rank algorithm to be used as a Semantic Web search engine. Relevance is measured as the probability of finding the connections made by the user at the time of the query, as well as the information contained in the base knowledge of the Semantic Web environment. By the use of "virtual links" between the concepts in a page, which are obtained from the knowledge base, we can connect concepts and components of a page and increase the probability score for a better ranking. By creating these connections, this study also looks to eliminate the possibility of getting results equal to zero, and to provide a tie-breaker solution when two or more pages obtain the same score.

IV. EXISTING SYSTEM

1. Within the literature, whereas there area unit some connected work, like net ranking spam detection, on-line review spam detection and mobile App recommendation, the matter of detective work ranking fraud for mobile Apps remains under-explored.[4]2. Typically speaking, the connected works of this paper are often sorted into 3 classes.

3. The primary class is regarding net ranking spam detection.

4. The second class is targeted on detective work on-line review spam.

5. Finally, the third class includes the studies on mobile App recommendation.

To the best of our knowledge, there is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud. Thus, we develop four intuitive baselines and invite human evaluators to validate the effectiveness of our approach.

Existing system do also have several limitations like;

Some of the existing approaches can be used for anomaly detection from historical rating and review records, they are not able to extract fraud evidences for a given time period (i.e., leading session).[6]

It is not able to detect ranking fraud happened in Apps' historical leading sessions

There is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud.[3]

Table 1: Comparative Analysis

PAPER	PROPOSED	ADVANTAGES	BASIC METHOD
Discovery of Ranking Fraud for Mobile Apps	System shows ranking fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records	Identify leading events and sessions by scanning historical ranking records only once	Optimization based aggregation method
Ranking Fraud Detection for Mobile Apps using Evidence Aggregation Method	Author proposed a ranking fraud detection process where there are some evidence considered and Integrated to yet an aggregated result which is must reliable.it happens in leading sessions	Reliable system	Evidence aggregation method
Review Spam Detection via Temporal Pattern Discovery	They propose a hierarchical detection criterion to detect SR spam attacks robustly and accurately. This feature is especially useful for online review website quality and trust monitoring.	This process continues until one reaches the desired resolution such that the time of SR spam attacks can be easily pinpointed.	Link spamicity
An unsupervised learning algorithm for rank aggregation	Author proposed a novel method to the rank aggregation problem by providing an optimization issues to discover a linear combination of ranking functions which exploits agreement	Effectiveness of the proposed technique is measure based on a data fusion task across ad hoc retrieval systems	Effectiveness of the proposed technique is measure based on a data fusion task across ad hoc retrieval systems
Detection of Ranking Fraud for Mobile App Using Fuzzy Logic	The pendulum algorithm decides whether the input of the Fuzzyfication come under truth value ranges between completely true or completely false	Extract the useful knowledge from huge amount of data	Pendulum Method

V. PROBLEM STATEMENT

The number of Apps has designed or increased at a vast speed across a couple of years. To refreshing the development of Apps many App stores introduced everyday basis Apps leaderboards chart, which shows the positions of nearly all famous Apps. A Top most rank on the chart generally lead to several downloads and earnings in million dollars, instead of depending on traditional marketing ways. Fake App creators apply some fraud full actions to intentionally improve their Apps and in the end inflate the chart positions on an App Store. This is normally done by utilizing “bot farms” or “human water armies”[8] to manipulate the App downloads rating and comments in an extremely limited time period. Although some of the existing approaches can be used for anomaly detection from historical rating and review records, this are not able to extract fraud evidences for a given time period (i.e., leading session)[1]. There is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud. This paper 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, this paper find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, these papers characterizes some fraud evidences from Apps’ historical ranking records, and develop three functions to extract such ranking based fraud evidences.[7]

VI. PROPOSED SYSTEM

There square measure 2 main steps for mining leading sessions. First, this paper ought to discover leading events from the App’s historical ranking records.

Second, this paper ought to merge adjacent leading events for constructing leading sessions. Specifically algorithmic program demonstrates the pseudo code of mining leading sessions for a given App In algorithmic program , they denote every leading event e and session s as tuples < begin, finish > and < begin, end,Es > severally, wherever metal is that the set of leading events in session s. Specifically, they initial extract individual leading event e for the given App a (i.e., Step two to 7) from the start time. for every extracted individual leading event e, they check the time span between e and also the current leading session s to make a decision whether or not they belong to an equivalent leading session.

Algorithm: **Forgery using Mathematical Evidences**

input 1: α is historical ranking record;
input 2: The ranking threshold
input 3: The merging threshold ;
output: The set of a’s leading sessions ;
Initialization: S α = **Ǿ**

1. For i=1;

2. s =<t^s_{start}; t^e_{end}; E_s >;
3. S_a U= s; s = Ø is a new session;
4. E_s = e; $t^s_{start} = t^e_{start}$; $t^s_{end} = t^e_{end}$;
5. t^e_{start} = 0; e = Ø is a new leading event;
6. return S_a
7. $t^e_{end} = t_{i-1}$; e =<t^e_{start}, t^e_{end}>;
8. if E_s == Ø then
9. E_s U= e; t^s_{start} = $t^s_{start} = t^e_{start}$; $t^s_{end} = t^e_{end}$;
10. else if ($t^e_{start} - t^s_{start}$) <Ø then
11. E_s U= e; $t^s_{end} = t^e_{end}$;
12. Define Rank, Rating, Review for an app.
13. Set counter for app ranking threshold = S α
14. If Rank > download threshold(Es) &
15. If Rating > Session Threshold(Sα) &
16. If Review metadata contains words like "Nice app", "Good App", etc.
17. Then App category is Fake.

VII. MATHEMETICAL MODEL

Two shape parameters $\theta 1$ and $\theta 2$ to quantify the ranking patterns of the rising phase and the recession phase of App a's leading event e, which can be computed by , **S** α = **Ø**
Here, this equation define a fraud signature θs for a leading session as follows, where |Es| is the number of leading events. [4]

Here, this paper propose to use the popular Gaussian approximation to compute the p-value with the above hypotheses. Specifically, this equation assume follows the Gaussian distribution, ~(,), where and can be learnt by the classic maximum-likelihood estimation (MLE) method from the observations of in all Apps' historical leading sessions.[5]

VIII. SYSTEM ARCHITECTURE

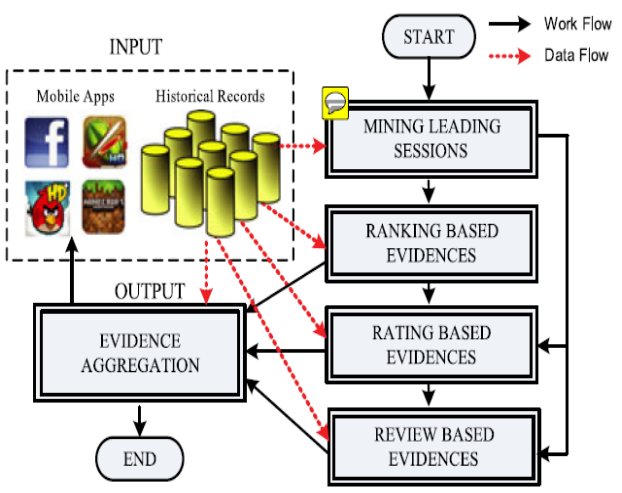

Fig 1: System Architecture

A. MINING LEADING SESSIONS

In the first module, this rule develops system surroundings with the small print of App like Associate in Nursing app store. Intuitively, the leading sessions of a mobile App represents its of recognition, that the ranking manipulation can solely happen in these leading sessions. Therefore, the matter of police investigation ranking fraud is to notice deceitful leading sessions. On this line, the primary task is a way to mine the leading session of a mobile App from its historical ranking records. These area unit 2 main steps for mining leading sessions. First, this paper ought to discover leading events from the App's historical ranking records.[6]

B. RANKING BASED EVIDENCES

In next module the system tries to find ranking fraud made in an app over the time. For this purpose we use the history of app to determine whether the ranking for the app has been modified according to recent trends or forged for publicity. To find the ranking fraud we check the ranking primarily based on no. of downloads for the app in a particular time period. If the downloads are made in a particular time period and not over the longer time then such apps are distinguished as forged apps.[2][1]

C. RATING BASED EVIDENCES

In the third module, we emphasis on development of such an mathematical module which can check on user login ratings. A user tends to provide rating based on performance of an app post its download. If any user has provided an rating without any download then it might go in forged rating, also we try to find lifetime threshold for an application and compare it with the current rating of an app. If the systems tries to cross this barrier multiple times in a short period than such an app is assumed to be forged.[1][2]

D. REVIEW BASED EVIDENCES

In fourth module, this algorithmic program add the Review primarily based Evidences module during this paper. Besides ratings, most of the App stores conjointly permit users to write down some matter comments as App reviews. Such reviews will mirror the non-public perceptions and usage experiences of existing users for explicit mobile Apps. Indeed, review manipulation is one in every of the foremost vital perspective of App ranking fraud. Specially, before downloading or getting a brand new mobile App, users typically initial scan its historical reviews to for his or her deciding, and a mobile App contains a lot of positive reviews might attract a lot of users to transfer. To find whether an app is real or fake we check the metadata content along with the app. This data provides the very important reviews made to an app. Fake reviews often tend to be small and precise without any information regarding to application. Actual reviews provides many kinds of solutions to make app better.

E. EVIDENCE AGGREGATION

In fifth module, this algorithmic program develops the proof Aggregation module to the present paper. Once extracting three sorts of fraud evidences, following challenges is a way

to mix them for ranking fraud detection. Indeed, there are several ranking and proof aggregation strategies within the literature, like permutation primarily based models score based models and Demster-Shafer rules. However, a number of these strategies target learning a worldwide ranking for all candidates. This is often not correct foe detective work ranking fraud for brand new Apps. Alternative strategies at supported supervised learning techniques, that rely upon the labeled coaching knowledge and erroneous to be exploited. Instead, this module propose associate degree unattended approach supported fraud similarity to mix these evidences.

IX. ADVANTAGES

1. The planned framework is ascendable and might be extended with alternative domain generated evidences for ranking fraud detection.

2. Experimental results show the effectiveness of the planned system, the measurability of the detection formula in addition as some regularity of ranking fraud activities.

3. To the simplest of our data, there's no existing benchmark to make a decision that leading sessions or Apps very contain ranking fraud. Thus, this paper develop four intuitive baselines and invite 5 human evaluators to validate the effectiveness of this paper approach proof Aggregation based mostly Ranking Fraud Detection.

4. Block the malware once the applying downloaded.

5. Transfer computer code supported risk score.

6. Application uses safely.

X. CONCLUSION

Paper has proposed an unique mobile application forgery detection system that relies on mining and leading session. Specifically, the paper tend to find forgery made within an app over its lifetime and provided a technique for mining leading sessions for every App from historical ranking records. Further in this paper we have defined the mathematical logic used to known ranking primarily based evidences, rating evidences and review evidences for sleuthing ranking fraud. Moreover, this paper has planned an optimization primarily based aggregation technique to integrate all the evidences for evaluating the believability of leading sessions from mobile Apps.

REFERENCES

[1] Hengshu Zhu, Hui Xiong, Senior Member, IEEE, Yong Ge, and Enhong Chen, Senior Member, IEEE L. Azzopardi, M. Girolami, and K. V. Risjbergen, "*Investigating the relationship between language model perplexity and ir precision- recall measures*," in Proc. 26th Int. Conf. Res. Develop. Inform. Retrieval, 2003, pp. 369–370.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "*Latent Dirichlet allocation*," J. Mach. Learn. Res., pp. 993–1022, 2003.

[3] Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "*A taxi driving fraud detection system*," in Proc. IEEE 11th Int. Conf. Data Mining, 2011, pp. 181–190.

[4] D. F. Gleich and L.-h. Lim, "*Rank aggregation via nuclear norm minimization*," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 60–68.

[5] L. Griffiths and M. Steyvers, "*Finding scientific topics*," Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.

[6] G. Heinrich, *Parameter estimation for text analysis*, "Univ. Leipzig, Leipzig, Germany, Tech.Rep., http://faculty.cs.byu.edu/~ringger/ CS601R/papers/Heinrich-GibbsLDA.pdf, 2008.

[7] N. Jindal and B. Liu, "*Opinion spam and analysis*," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219–230.

[8] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li, "*Supervised rank aggregation*," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 481–490.

