© 2017 IJSRST | Volume 3 | Issue 1 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Discovery of Ranking Fraud Using Evidence Aggregation Approach for Mobile Apps

Katkade Roshani, Ghane Deepali, Nimse Megha, Jangam Nandini

Department of Computer Engineering, MCERC, Nashik, Maharashtra, India

ABSTRACT

Nowadays everyone is using smart phone. There is need of various applications to be installed on smart phone. To download application smart phone user have to visit Apps store such as Google Play Store, Apples store etc. When user visit play store then he or she is able to see the various applications list. This list is built on the basis of promotions and advertisements. User doesn't have knowledge about the application (i.e. Whether the application is useful or not.) So user looks at the list and downloads the applications mostly from first page of App Store. But sometimes it happens that the downloaded application won t work or is useless. That means it is fraud in mobile application list. To avoid this fraud, there is a need to develop an application which will list out the useful applications. To list the applications first we are going to find the active period of the application named as leading session. We are also investing the three types of evidences: 1. Ranking based evidence. 2. Rating based evidence. 3. Review based evidence. Using these three evidences finally we are calculating aggregation of these evidences. We Evaluate our application with real world data collected form play store for long time period.

Keywords: Mobile Apps, ranking fraud detection, evidence aggregation, historical ranking records, rating and review

I. INTRODUCTION

In all over the world for the mobile electronic devices are a very vast collection of millions of mobile Apps. These Apps developed by App Developer and Post at leaderboard for ranking purpose. The number of versatile mobile Apps has developed at a stunning speed in the course of recent years. For instance, at the month end of April 2013, there are 1.6 million and more than those apps at Apple's App store and Google Play. To fortify the advancement of portable Apps, numerous App stores introduced day by day App leader boards, which show the graph positions of almost all well-known Apps. In fact, the App leader board is a standout amongst the most essential routes for advancing portable Apps. A top most position on the leader board more popular is

the app is the fact. Top ranked app have more amount of downloads and earnings in million dollars. In this form, App designers have a tendency to investigate different ways for getting the higher position in leader board for example, promoting advertisement for their Apps keeping in mind the end goal to have their Apps ranked as top rank as possible in Application leaderboard.

Ranking fraud in the mobile app market refers to fraud activities which have a purpose of increasing apps in list. Indeed, it becomes more common for app developers to use shady means, such as inflating their apps' sales or posting fraud App ratings, to commit ranking fraud. While the importance of detect ranking fraud has been widely recognized, there is limited understanding and research in this area. To this end, in this paper, we provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile apps. Specifically, we first propose to accurately locate the ranking fraud by mining the active periods, namely leading sessions, of mobile of global anomaly of app ranking. In Rating Based Evidences, specifically, after an App has been published, it can be rated by any user who downloaded it. Indeed, user rating is one of the most important features of App advertisement. An App which has higher rating may attract more users to download and can also be ranked higher in the leaderboard. Especially, this paper proposes a simple and effective algorithm to recognize the leading sessions of each mobile App based on its historical ranking records. This is one of the fraud evidence. Also, rating and review history, which gives some anomaly patterns from apps historical rating and reviews records. 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. First propose an effective algorithm to identify the leading sessions of each App based on its historical ranking records. Then, with the analysis of Apps' ranking behaviors, find out the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps.

The organization of this document is as follows. In Section 2 (Literature Survey),give details of existing methods. Section 3 give detail of any modifications to equipment or equipment constructed specifically for the study and, if pertinent, provide illustrations of the modifications. In Section 3 (Result and Discussion), present research findings and analysis of those findings. Discussed in Section 4 concludes the paper.

II. METHODS AND MATERIAL

1. Literature Survey

Hui Xiong[1] discovered ranking fraud detection System for mobile Apps but it is still under study research. To fill this crucial lack, we propose to develop a ranking fraud detection system for mobile Apps. We also determine several important challenges. First challenge, in the whole life cycle of an App, the ranking fraud does not always happen, so we need to detect the time when fraud happens. Finally, due to the dynamic nature of chart rankings, it is difficult to find and verify the evidences associated with ranking fraud, which motivates us to discover some implicit fraud patterns of mobile Apps as evidences.

D. M. Blei[3] has proposed latent dirichelet allocation of generative probabilistic model for Accumulations of discrete data such as text corpora. LDA is a three-level various leveled Bayesian model, in which everything of an accumulation is displayed as a limited blend over an basic set of points. Every subject is, thus, demonstrated as a boundless blend over a basic arrangement of point probabilities. In the context of text modeling, the topic probabilities give an explicit representation of a document. We introduced efficient approximate inference techniques based on variation processes and an EM algorithm for empirical Bayes parameter estimation. We report results in record modeling, text classification, and collaborative filtering, comparing to a mixture of unigrams model and the Probabilistic LSI model.

Y. Ge, H. Xiong[2] has proposed taxi driving fraud detection system for Advances in GPS following innovation have enabled us to introduce GPS beacons in city taxis to gather a lot of GPS follows under operational time requirements. These GPS follows give unparalleled chances to us to reveal taxi driving extortion exercises. In this paper, add to a taxi driving misrepresentation identification framework, which can efficiently explore taxi driving extortion. In this framework, first give capacities to discover two parts of confirmations: travel course proof and driving separation proof. Besides, a third capacity is intended to consolidate the two parts of proofs in view of Dempster-Shafer hypothesis. To actualize the framework, first recognize introducing destinations from a lot of taxi GPS logs.

A. Klementiev, D. Roth[7] An unsupervised learning algorithm for rank aggregation Many applications in data recovery, natural language processing, information mining, and related fields require a positioning of cases regarding indicated criteria instead of a grouping. Moreover, for some such issues, various set up positioning models have been all around concentrated on and it is alluring to consolidate their outcomes into a joint positioning, formalism indicated as rank accumulation. This paper exhibits a novel unsupervised learning algorithm for rank accumulation (ULARA) which gives back a direct blend of the individual positioning capacities of ranking functions in view of the standard of compensating requesting understanding between the rankers.

D. F. Gleich[15] has proposed Rank aggregation via nuclear norm minimization with the method of rank aggregation is informally interwoven with the structure of skew-symmetric matrices and applies recent approach in the theory and algorithms of matrix completion to skew-symmetric matrices. This mix of thoughts delivers another system for positioning an arrangement of things. The need of our plan is that a rank aggregation shows a partially filled skew-symmetric matrix. Here extend an algorithm for matrix completion to hold skew-symmetric information and utilize that to take out ranks for each item. This algorithm applies to both pairwise comparison and rating data. Because it is based on matrix completion, it is vigorous to both noise and inadequate information.

Klementiev, D. Roth, and K. Small[8] has proposed Unsupervised rank aggregation with distance-based models which needs to incorporate the arrangement of rankings regularly manages collecting and it just comes up when a specific positioned information is produced. Despite the fact that the different heuristic and managed learning ways to deal with rank total, a prerequisite of area information and directed positioned information exists. Along these lines, to resolve this issue, a structure is proposed for learning total rankings without supervision. This system is instantiated for the instances of permutations and combinations of top-k records.

E. P. Lim [10] has proposed product review spammer detection to locate customers producing unsolicited mail reviews or evaluation spammers. He became aware of several feature behaviors of evaluate spammers and model those behaviors so as to discover the spammers. Particularly, they seek to model the following behaviors. First, spammers might also target unique merchandise or product organizations to be able to maximize their effect. They tend to deviate from the alternative reviewer of their rankings of products. They advocate scoring methods to degree the degree of unsolicited mail for every reviewer and apply them on an Amazon overview data set then pick a sub-set of notably suspicious reviewers for further scrutiny by using our user evaluates with the help of a web primarily based spammer evaluation software program particularly evolved for person assessment experiments.

2. Proposed System

1. System Architecture

This section gives overview of framework of system Ranking fraud usually happens in these leading sessions. After careful understanding the system has been divided into the three evidences and at last aggregate these evidences as a result. There are two main phases of this system.

i)Identifying the leading sessions for mobile apps.
ii) Identifying evidences for ranking fraud detection

Leading Events

(Leading Event)-Given a positioning limit a main occasion of App contains a time period range. Relating rankings of mobile App a, Note that we apply a positioning edge which is normally smaller than K , here on the grounds that K may be huge, and the positioning records past are not exceptionally helpful for recognizing the positioning controls. Moreover, we additionally find that a few Apps have a few nearby driving events. Especially, a main occasion which does not have other close-by neighbors can likewise be dealt with as an uncommon driving session.

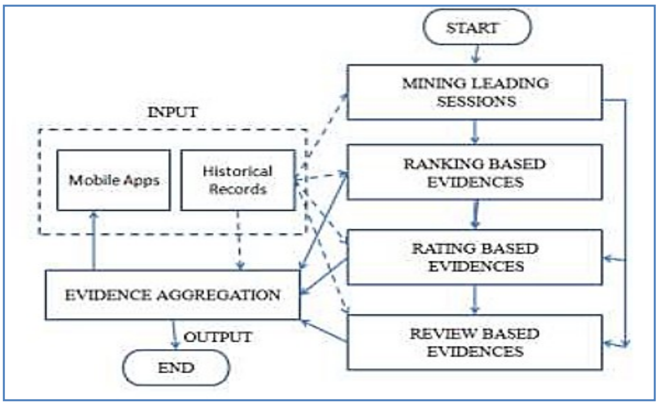

Figure 1. System Architecture

Leading Sessions

A main session of App contains a period range Ts and n adjoining driving occasion. Application speak to its times of fame, so the positioning control will just occur in these driving sessions. Along these lines, the issue of recognizing positioning extortion is to distinguish fake driving sessions. Along this line, the first assignment is the means by which to mine the main sessions of a horde. Leading session is calculated from closable leading event.

Identifying the Leading Sessions for Mobile APPs:

In mining leading session algorithm there are two important steps for calculating particular period in which fraud is happened for that particular mobile app. The first step is to search leading events from the mobile App's historical ranking records and second is for merging neighbouring leading events for developing leading sessions. Specifically, firstly extract individual leading event e for the given App from the starting time. For each and every extracted individual leading event e, we check the time span between e and the current leading session s to decide whether they include to the similar leading session. Thus, this algorithm can identify leading events and sessions by scanning mobile app a’s historical records only once.

Identifying evidences for ranking fraud detection:

Identifying different evidences for ranking fraud detection is applied on output of mining leading session algorithm. Step by step three evidences applied are ranking, rating and review based.

- **Rating Based Evidences:** In Rating Based Evidences, specifically, after an App has been published, it can be rated by any user who downloaded it. Indeed, user rating is one of the most important features of App advertisement. An App which has higher rating may attract more users to download.
- **Review Based Evidences:** Besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such review Can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Specifically, before to download. Therefore, imposters often post fake reviews in the leading sessions of a specific App in order to inflate the App download, and thus propel the App’s ranking position in the leaderboard.
- **Ranking Based Evidences:** By investigating the Apps' historical positioning records, we watch that Apps' positioning practices in a main occasion dependably fulfill a particular positioning example, which comprises of three diverse positioning stages, to be specific ,rising stage, keeping up stage and retreat stage. In particular, in every driving occasion, an App's positioning first increments to a top position in the leaderboard (i.e., rising stage), then keeps such top position for a period (i.e., looking after stage), lastly diminishes till the end of the occasion (i.e., retreat stage. Can also be ranked higher in the leader board. Thus, rating manipulation is also an important perspective of ranking fraud. downloading a new mobile App, users often first read its historical reviews to ease their decision making, and a mobile App contains more positive reviews may attract more users.

Algorithm Used

There are two main steps for mining leading sessions. First, we need to discover leading events from the App’s historical ranking records. Second, we need to merge adjacent leading events for constructing leading sessions.

The Pseudo code of mining leading sessions for a given App are as follow.

Algorithm : Mining Leading Sessions

Input 1: ɑ’s Historical Ranking Records Rɑ;
Input 2: The ranking Threshold K*;
Input 3: The Merging threshold ∅;
Output: The set of a’s Leading sessions Sɑ;
Initialization :Sɑ= ∅;

1. Es=∅; e=∅; s=∅;t^estart=0;
2. **for each** i ∈[1,|Rɑ|]**do**
3. if $r^{\alpha}_i \leq K^*$ **and** t^estart==0 then
4. t^estart=t_i;
5. **else if** $r^{\alpha}_i > K^*$ **and** t^estart ≠ 0 **then**
6. // **Found one Event;**
7. t^eend=t_i-1; e=< t^estart , t^eend>
8. **if** Es **==**∅ **then**
9. Es ∪=e; $\mathbf{t}^s$start = t^estart; $\mathbf{t}^s$end= t^eend;
10. **else if** (t^estart - t^eend)< ∅ **then**
11. Es ∪=e; $\mathbf{t}^s$end= t^eend;
12. **else then**
13. //**found one session;**
14. s =< $\mathbf{t}^s$start , $\mathbf{t}^s$end ,Es >;
15. Sɑ ∪=s ; s=∅ is a new session;
16. Es={e} **;** $\mathbf{t}^s$start = t^estart; $\mathbf{t}^s$end= t^eend;
17. t^estart=0; e=∅ is a new leading event;

Return Sɑ

Algorithm: **Evidence Aggregation**

1. Analyze the historical records of mobile apps.
2. Differentiate the evidences as Ranking based, Rating based, Review based.
3. Aggregate these evidences.
4. Design Android application framework.

III. RESULTS AND DISCUSSION

Here, main attention is on extracting different evidences such as reviews, ratings, ranking and download information from historical records of data set. Data set contains the historical reviews, ratings of mobile apps. In the result parts calculates and merge the evidences with help of evidence aggregation method.

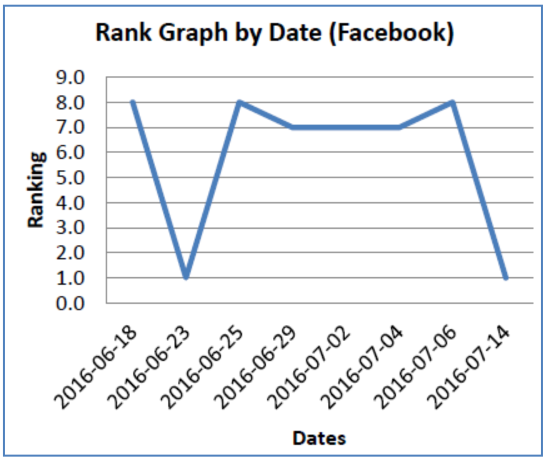

Figure 2. Rank Graph By Date For Fraud App.

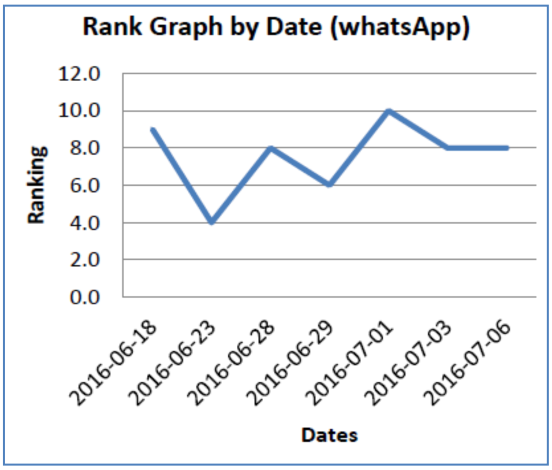

Figure 3. Rank Graph By Date For Normal App.

IV. CONCLUSION

In this paper, we develop ranking fraud detection system for mobile apps. It reviews various existing strategies used for internet or web spam detection, which is associated with the rating fraud for mobile Apps. Also, we've seen references for online review unsolicited mail detection and mobile App advice. By using mining the main sessions of mobile Apps, we aim to locate the ranking fraud. The leading classes works for detecting the nearby anomaly of App ratings. The machine targets to locate the ranking frauds based on three styles of evidences, including rating based evidences, ranking based evidences and comment based evidences. In addition, an optimization based totally aggregation method combines all of the three evidences to hit upon the fraud.

V. REFERENCES

[1] Hengshu Zhu, Hui Xiong Discovery of Ranking Fraud for Mobile Apps. IEEE TRANSACTIONS ON KNOWLEDGE ANDDATA ENGINEERING,2013.

[2] 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.

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

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

[5] 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.

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

[7] A. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach. Learn., 2007, pp. 616–623.

[8] A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with distance-based models," in Proc. 25th Int. Conf. Mach. Learn., 2008, pp. 472–479.

[9] A. Klementiev, D. Roth, K. Small, and I. Titov, "Unsupervised rank aggregation with domain-specific expertise," in Proc. 21st Int. Joint Conf. Artif. Intell., 2009, pp. 1101–1106.

[10] E.-P. Lim,V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers use rating behaviors," in Proc.19thACMInt. Conf. Inform. Knowl. Manage., 2010, pp. 939–948.

[11] 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.

[12] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, “Detecting spam web pages through content analysis,” in Proc. 15th Int. Conf. World Wide Web, 2006, pp. 83–92.
[13] K. Shi and K. Ali, “Getjar mobile application recommendations with very sparse datasets,” in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 204–212.
[14] N. Spirin and J. Han, “Survey on web spam detection: Principles and algorithms,” SIGKDD Explor. Newslett., vol. 13, no. 2, pp. 50–64, May 2012.
[15] 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.