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 6, June 2017

Survey On: "To Detect Ranking Fraud App with the Help of Discovery of Historical Record"

Pritam Porate, M. S. Nimbarte

M. Tech Student, Department of Computer Science and Engineering, Bapurao Deshmukh College of Engineering, Wardha, India

Professor, Department of Computer Science and Engineering, Bapurao Deshmukh College of Engineering, Wardha, India

ABSTRACT: The Mobile App is a very popular and well known concept due to the rapid advancement in the mobile technology. Due to the large number of mobile Apps, ranking fraud is the key challenge in front of the mobile App market. Ranking fraud refers to fraudulent or vulnerable activities which have a purpose of bumping up the Apps in the popularity list. While the importance and necessity of preventing ranking fraud has been widely recognized. In the existing system the leading event and leading session of an app is identified from the collected historical records. Then three different types of evidences are collected from the user feedbacks namely ranking based evidence, rating based evidence and review based evidence. These three evidences are aggregated by using evidence aggregation method. In the proposed system additionally, we are proposing two enhancements. Firstly, we are using Approval of scores by the admin to identify the exact reviews and rating scores. Secondly, the fake feedbacks by a same person for pushing up that app on the leader board are restricted. Two different constraints are considered for accepting the feedback given to an application. The first constraint is that an app can be rated only once from a user login and the second is implemented with the aid of IP address that limits the number of user login logged per day. Finally, the proposed system will be evaluated with real-world App data which is to be collected from the App Store for a long time period. The first constraint is that an app can be rated only once from a user login. And the second is implemented with the aid of MAC address

KEYWORDS: Mobile Apps, Ranking Fraud Detection, Evidence Aggregation, Historical Ranking Records, Rating and Review, Recommendation app.

I. INTRODUCTION

Ranking fraud in the mobile app market refers to fraudulent or deceptive activities which have a purpose of bumping up the apps in the popularity list. Indeed, it becomes more and more frequent for app developers to use shady means, such as inflating their apps' sales or posting phony App ratings, to commit ranking fraud. While the importance of preventing 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 Apps. Such leading sessions can be leveraged for detecting the local anomaly

DOI: 10.15680/IJIRCCE.2017. 0506027

instead of global anomaly of app rankings. Furthermore, we investigate 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 statistical hypotheses tests. 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 leader board. Thus, rating manipulation is also an important perspective of ranking fraud. In Review Based Evidences, besides ratings, most of the App stores also allow users to write some textual comments as App reviews. 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.

II. LITERATURE REVIEW

a. A Comprehensive Study Of A Mathematical And Algorithmic Framework :

Author Klementiev, D. Roth, and K. Small suggest that explains the need to meaningfully combine sets of rankings often comes up when one deals with ranked data. Although a number of heuristic and supervised learning approaches to rank aggregation exist, they require domain knowledge or supervised ranked data, both of which are expensive to acquire. In order to address these limitations, they propose a mathematical and algorithmic framework for learning to aggregate (partial) rankings without supervision. The framework for the cases of combining permutations and combining top-k lists, and propose a novel metric for the latter. Experiments in both scenarios demonstrate the effectiveness of the proposed formalism.

b. A Survey of Web Spam Detection Techniques :

Author Ee-Peng Lim et al. presented a number of detecting Product Review Spammers using Rating Behaviors to detect users generating spam reviews or review spammers. We identify several characteristic behaviors of review spammers and model these behaviors so as to detect the spammers. Now a days, with reference to increasing info in net, search engines are thought of as a tool to enter the net. Then gift an inventory of results associated with user question. A legal thanks to increase sites rank within the list results of search engines is increasing the standard of web sites pages, however this technique is time consuming and expensive. Another technique is use outlaw and unethical ways to extend the rank in search engines. The hassle of deceiving search engines is named net spam. Web spam has been through of collectively of the common issues in search ngines, and it's been planed once search engines appeared for the primary time. The aim of net spam is to vary the page rank in question results, during this means, it's placed in an exceedingly rank beyond traditional conditions, and it's ideally placed among ten prime sites of question leads to varied queries.

c. HySAD: A Semi-Supervised Hybrid Shilling Attack Detector for Trustworthy Product :

Author Ee-Peng shows that Shilling attackers apply biased rating profiles to recommender systems for manipulating online product recommendations. Though several studies are dedicated to shilling attack detection, few of them will handle the hybrid shilling attacks that sometimes happen in follow, and therefore the studies for real life applications area unit seldom seen. Morever, very little attention profiles, though there area unit typically a number of labeled however various unlabeled users accessible in follow. This paper presents a Hybrid Shilling Attack Detector, or HySAD for brief, to tackle these issues. Above all, HySAD introduces MC-Relief to pick effective detection metrics, and Semi-supervised Naïve mathematician to exactly separate Random-Filler model attackers and Average-Filler model attackers from traditional users.

d. **A Semantic Association Page Rank Algorithm for Web Search Engines :**

Author Manuel Rojas suggest that in this paper propose a relation-based page rank formula to be used as a Semantic Web search engine. connectedness is measured is because the likely food of finding the connections created by the user at the time of the questions, yet because the information contained within the base information of the Semantic Web environment. By the employment of “virtual links” between the ideas in a page, that area unit obtained from the knowledge base, connect this paper 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 answer when two or more pages acquire the same score. This paper are able to connect idea and parts of page and increase the likelihood score for a more robust ranking for mobile apps

1. The primary class is regarding net ranking spam detection.
2. The second class is targeted on detective work on-line review spam.
3. Finally, the third class includes the studies on mobile App recommendation.

e. A Taxi Driving Fraud Detection System :

Author Ge, H. Xiong, C. Liu, and Z.-H. Zhou explains the advances in GPS tracking technology have enabled us to install GPS tracking devices in city taxis to collect a large amount of GPS traces under operational time constraints. These GPS traces provide unparallel opportunities for us to uncover taxi driving fraud activities. In this paper, the author developed a taxi driving fraud detection system, which is able to systematically investigate taxi driving fraud. In this system, we first provide functions to find two aspects of evidences: travel route evidence and driving distance evidence. Furthermore,a third function is designed to combine the two aspects of evidences based on Dempster-Shafer theory. To implement the system, we first identify interesting sites from a large amount of taxi GPS logs. Then, a parameter-free method to mine the travel route evidences. Also the introduction of route mark is used to represent a typical driving path from an interesting site to another one.

f. Opinion spam and analysis in Web Search Data Mining :

Author N. Jindal and B. Liu suggest that In this paper, be study this issue in the context of product reviews, which are opinion rich and are broadly used by clients and product manufacturers. In the past two years, several startup companies also appeared which collective opinions from product reviews. It is thus high time to study spam in reviews. To the best of our knowledge, there is still no published study on this topic, although Web spam and email spam have been investigated expansively. In this will see that opinion spam is somewhat different from network spam and email spam, and thus requires different detection techniques. Based on the analysis of 5.8 million reviews and 2.14 million reviewers from amazon.com, in this show that opinion spam in reviews is widespread. This paper analyzes such spam activities and presents some fresh techniques to detect them.

g. A System For Directing Supervised Rank Aggregation.

Author Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li shows In this paper we first give a general system for directing Supervised Rank Aggregation. We demonstrate that we can characterize directed learning techniques relating to the current unsupervised strategies, for example, Board Count and Markov Chain based routines by abusing the system. At that point we predominantly research the administered forms of Markov Chain based techniques in this paper, in light of the fact that past work demonstrates that their unsupervised partners are unrivaled. Things being what they are turns out, on the other hand, that the streamlining issues for the Markov Chain based routines are hard, in light of the fact that they are not curved improvement issues. We have the capacity to add to a system the enhancement of one Markov Chain based technique, called Supervised MC2.Specifically, we demonstrate that we can change the advancement issue into that of Semi positive Programming.

h. Discovery Of Ranking Fraud For Mobile Apps :

Author Hengshu Zhu,Hui Xiong,,Yong Ge,and Enhong Chen suggest that this paper developed a ranking fraud system for mobile Apps. Specifically, This paper tend to initial showed that ranking fraud happened in leading sessions and provided a technique for mining leading sessions for every App from historical ranking records. Then, This paper tend to known ranking primarily based evidences, rating evidences and review evidences for sleuthing ranking fraud. Moreover, This paper tend to planned AN optimization primarily based aggregation technique to integrate all the evidences for evaluating the believability of leading sessions from mobile Apps.

III. PROPOSED WORK

In proposed system we overcome the problems of Mining leading session algorithm which is based on ranking, review & rating. Detection of ranking fraud for mobile Apps is still under a subject to research. To fill this crucial lack, we propose to develop a ranking fraud detection system for Apps. We also determine several important phases. First phase, 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. This phase can be considered as detecting the local anomaly in place of global anomaly of mobile Apps. Second phase, it is important to have a scalable way to positively detect ranking fraud without using any basis information, as there are huge number of mobile Apps, it is very difficult to manually label ranking fraud for each App. 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. The users who are newly logging to the app stores, that time users details are stored in the database with the MAC Address of system. If users regularly download the his app and gives fake rating and reviews then admin block his account from play store and next time cannot allow him to upload the app at play store. This all doing with the help of historical record and MAC Address. In recent activities duplicate version of an application not burned or blocked. This is the major defect. Higher rank leads huge number of downloads and the app developer will get more profit. In this they allow Fake Application also. User not understanding the Fake Apps then the user also gives the reviews in the fake application. Exact Review or Ratings or Ranking Percentage are not correctly Calculated. In this paper we introduce admin to manage the ranking evidence to minimize the arrival of fake apps, and then the rating and reviews are correctly calculated. Also in this used KNN, K-means and AES algorithm for the finding the fraud app. If user continuously doing fraud then we can block this fraud developer and prevent the AppStore.

EVIDENCE AGGREGATION ALGORITHM

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

Step1: Analyzing of historical records is nothing but obtaining the app related information from Google play store and apple store. Historical records consist Rank of the applications in leaderboard, Rating given by users to apps, different reviews of different types of users, no of downloads off the apps.

a. Ranking Based Evidences

The ranking is depend on the rating. If user gives the more and more rating to the any app then the ranking is automatically increased in AppStore. By analyzing the Apps' historical ranking records, we observe that Apps' ranking behaviors' in a leading event always satisfy a particular ranking pattern, which include the three dissimilar ranking phases, namely, rising phase, maintaining phase and recession phase. Particularly, in every one leading event, an App's ranking first increases to a peak position in the leaderboard (i.e., rising phase), at that time keeps such peak position for a period (i.e., maintaining phase), and lastly lessen till the end of the event (i.e., recession phase). Fig. 1 shows an example of different ranking phases of a leading event. In addition, such a ranking pattern shows a fundamental understanding of leading event. In the following, we formally define the three ranking phases of a leading event.

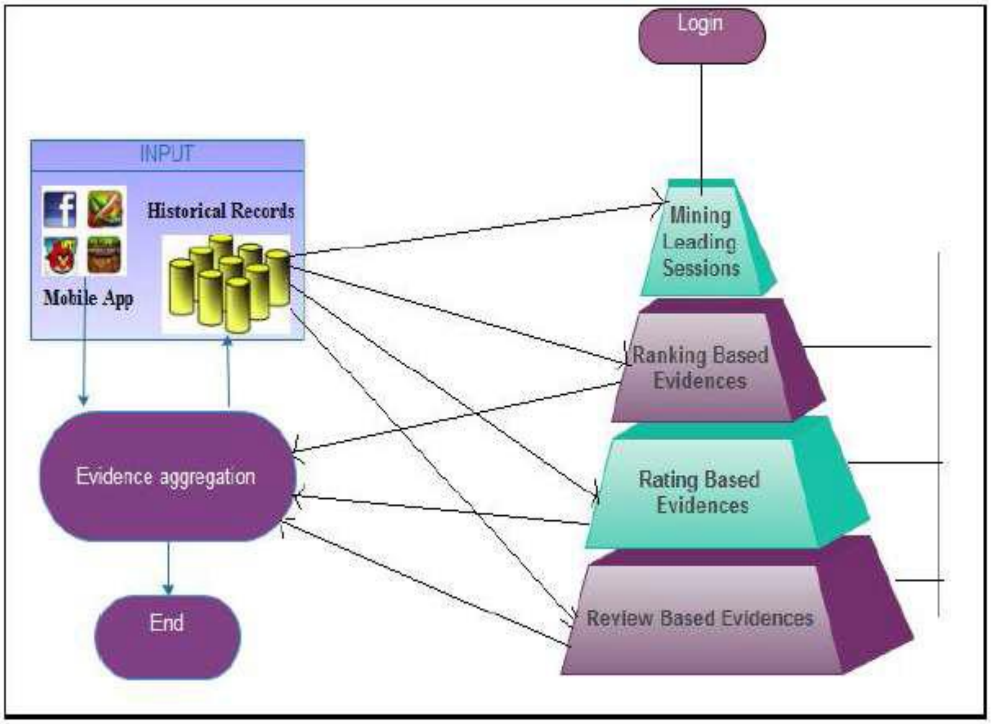

Fig. 1: System Model

b. Rating Based Evidences

For ranking fraud detection are uses the ranking based evidences. However, sometimes, it is not sufficient to only use ranking based evidences. For instance, some Apps developed by the famous developers, such as Game loft, may have some leading events with large values of u1 due to the developers' credibility and the "word-of-mouth" advertising effect. Additionally, some of the legal marketing services, such as "limited-time discount", may also result in significant ranking based evidences. To solve that problem, we additionally study how to extract fraud evidences from Apps' historical rating records. 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 valuable features of App advertisement. An App which has higher rating may attract more users to download and also can gives ranked higher in the leaderboard. Thus, rating manipulation is also a valuable perspective of ranking fraud. Innocently, if an App has ranking fraud in a leading session s, the ratings during the time period of s may have inconsistency patterns merged with its historical ratings, which can be used for constructing rating based evidences.

c. Review Based Evidences

Review is nothing but the feedback or comments which are given by the user .With the help of review user can share his experience with the App. It may be bad or good review but it is important. In addition ratings, most of the App stores also permit users to write some textual comments as App reviews. Such reviews can indicates the individual perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most valuable perspective of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users usually first read its historical reviews to ease their decision making, and a mobile App contains more encouraging reviews may captivate more users to download. Therefore, imposters often post fake reviews in the leading sessions of a particular App in order to increases the App downloads, and thus propel the App's ranking position in the leaderboard. For all that previous works on review spam detection have been reported in recent years the issue of detecting the local inconsistency of reviews in the leading sessions and capturing them as evidences for ranking fraud detection are still under explored. For this purpose, here we propose two fraud evidences for detecting ranking fraud based on Apps' review behaviours in leading sessions.

d. Evidence Aggregation

After extracting all three types of fraud evidences, then the next challenge is how to combine them for ranking fraud detection. In addition, there are many methods of ranking and evidence aggregation in the literature, such as permutation based models score based models and Dempster Shafer rules. However, some of these methods focus on learning a global ranking for all applicants. This way is not proper for detecting ranking fraud for new Apps. Distinct methods are based on supervised learning techniques, which rely on the labelled training data and are hard to be exploited. Rather, we suggest an unsupervised approach based on fraud similarity to combine these evidences.

Detecting ranking fraud of mobile Apps is actually to detect ranking fraud within leading sessions of mobile Apps. Specifically, we first propose 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, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, we characterize some fraud evidences from Apps' historical ranking records, and develop three functions to extract such ranking based fraud evidences.

IV. CONCLUSIONS

We conclude that, to develop a ranking fraud detection system for mobile Apps. we first discover that ranking fraud occur in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. In that case, we identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. Furthermore, we proposed an optimization based aggregation method to integrate all the evidences for evaluating the reliability of leading sessions from mobile Apps. That all the evidences can be modeled by statistical hypothesis tests for the unique perspective of this approach, thus it is easy to be extended with other evidences from domain knowledge to detect ranking fraud. Ultimately, we validate the proposed system with extensive experiments on real world App data collected from the google play store. Experimental results showed the effectiveness of the proposed approach.

REFERENCES

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

[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, pp. 181– 190, 2011.

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

[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]. B. Zhou, J. Pei, and Z. Tang. “A spamicity approach to web spam detection”. In *Proceedings of the 2008 SIAM International Conference on Data Mining*, SDM’08, pages 277–288, 2008.

[7]. A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly. “Detecting spam web pages through content analysis”. In *Proceedings of the 15th international conference on World Wide Web*, WWW ’06, pages 83–92, 2006.

[8]. N. Spirin and J. Han. “Survey on web spam detection: principles and algorithms”. *SIGKDD Explor. Newsl.*, 13(2):50–64, May 2012.

[9]. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. “Detecting product review spammers using rating behaviors”. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM ’10, pages 939–948, 2010.

[10]. Z.Wu, J.Wu, J. Cao, and D. Tao. “Hysad: a semi- supervised hybrid shilling attack detector for trustworthy product recommendation”. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD ’12, pages 985–993, 2012

[11]. S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD ’12, pages 823–831, 2012.

[12]. B. Yan and G. Chen. Appjoy: personalized mobile application discovery. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, MobiSys 11, pages 113– 126, 2011.

[13]. K. Shi and K. Ali. Getjar mobile application recommendations with very sparse datasets. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD ’12, pages 204–212, 2012.

[14]. H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian. Mining personal context-aware preferences for mobile users. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 1212–1217, 2012.

[15]. Hengshu Zhu, Hui XiongDiscovery of Ranking Fraud for Mobile Apps. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,2013.