International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 20 Issue 2 – FEBRUARY 2016.

ENHANCED SECURE ONLINE REPUTATION DEFENSE SYSTEM FROM FRAUD RATING USING ANOMALY DETECTION FOR MOBILE APPS

Mrs. K. Manohari, M.C.A, M.Phil[1] , Miss. C. Azhagurani [2]

[1] *Assistant Professor, Department of Computer Science, Theivanai Ammal College for Women (Autonomous), Villupuram-645 401, India. Email Id:* ***manohema@gmail.com***

[2] *M.Phil Scholar, Department of Computer Science, Theivanai Ammal College for Women (Autonomous), Villupuram-645 401, India. Email Id:* ***azhagurani.tvn @gmail.com***

***Abstract*-Ranking fraud in the mobile App business alludes to false or tricky exercises which have a motivation behind, knocking up the Apps in the fame list. To be sure, it turns out to be more incessant for App designers to utilize shady means, for example, expanding their Apps' business or posting imposter App evaluations, to confer positioning misrepresentation. While the significance of avoiding Ranking fraud has been generally perceived, there is constrained comprehension and examination here. This paper gives a holistic perspective of positioning misrepresentation and propose a Ranking fraud identification framework for mobile Apps. In particular, it is proposed to precisely find the mining, so as to pose extortion the dynamic periods, to be specific driving sessions of Mobile Apps. At last, assessment of the proposed framework is done with certifiable App information gathered from the Android market. In the investigations, this paper accepts the adequacy of the proposed framework, and demonstrates the identification's versatility calculation and also some consistency of positioning misrepresentation exercise.**

Keywords: Fraud ranking, Mobile Apps, ranking records, Rating and Review

I. INTRODUCTION

Web spam refers to all forms of malicious manipulation of user generated data so as to impudence usage patterns of the data. The number of mobile Apps has grown at a breath Taking rate over the past few years. For example, as of the end of April 2013, there are more than 1.6million Apps at Apple's App store and Google Play. To stimulate the development of mobile Apps, many App stores launched daily App leader boards, which demonstrate the chart rankings of most popular Apps. Indeed, the App leader boar's one of the most important way for promoting mobile Apps. A higher rank on the leader board usually leads to huge number of downloads and million dollars in the revenue. Therefore, App developers tend to explore various ways such as advertising campaigns to promote their Apps in order to have their Apps ranked as high as possible in such App leader boards. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, hence within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow. Indeed, our careful observation reveals that mobile Apps are not always ranked high in the leader board, but only in some leading events, which form different leading sessions. Note that we will introduce both leading events Ease of Use and leading sessions in detail later.

In other words, ranking fraud usually happens in these leading sessions. Therefore, 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 have characterized some fraud evidences from Apps' historical ranking records, and develop three functions to extract such ranking based fraud evidences. 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 f 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.

In addition, we develop an unsupervised evidence aggregation method to integrate these three types of evidences for evaluating the credibility of leading sessions from mobile Apps. It is worth noting that all the evidences

are extracted by modeling Apps' ranking, rating and review behaviors through statistical hypotheses tests. The propose framework is scalable and can be extended with other domain generate d evidences for ranking fraud detection. Finally, we evaluate the proposed system with real-world App data collected from the Apple's App s tore for a long time period, i.e., more than two years. Experimental results show the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity of ranking fraud activities. According to the definitions introduced in, a leading session is composed of several leading events. Therefore, we should first analyze the basic characteristics of leading evens for extracting fraud evidences. By the analyzing the Apps' historical ranking records, we observe those that Apps' ranking behaviors in a leading even always satisfy a specific ranking pattern, which consists of the three different ranking phase, namely, rising phase, maintaining phase and recession phase.

II. RELATED WORK

D. M. Blei, A. Y. Ng, and M. I. Jordan, introduces a unique model called as Dirichlet allocation (LDA) a generative probabilistic model for collections of discrete data such as text amount. Basically it is a three level hierarchical Bayesian model in which each element of a group is demonstrated as a finite mixture over a fundamental set of topics. Each topic is demonstrated as an infinite mixture over fundamental set of topic probabilities. With the reference of text modeling, the topic probabilities provide an open representation of a document. An efficient approximation inference technique is presented based on various methods and an EM algorithm for empirical Bayes parameter estimation is also presented. The results are reported in document Modeling, text classification and collaborative filtering, which compares to a collection of unigrams and probabilistic LSI model.

Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou , illustrated that growth in the field of GPS tracking technology have allowed the users to install GPS tracking devices in taxies to gather huge amount of GPS traces under some time period. These traces by GPS offered an unparalleled opportunity to uncover taxi driving fraud traces and then fraud detection system is proposed which is able to identify taxi driving fraud. First, two sort of function are uncovered here i.e. travel route evidence and driving distance evidence. Even a third function is developed to combine the previous functions based on Dempster-Shafer theory. First identification of interesting locations is done from tremendous amount of taxi GPS logs and then parameter free method is proposed to extract the travel route evidences. Secondly, concept of route mark is +++developed to illustrate the driving path between locations and based on those mark, specific model is characterized for the distribution of driving distance and discover the driving distance evidences. And finally, taxi driving fraud detection system with a large scale real world taxi GPS logs.

T. L. Griffiths and M. Steyvers, introduces the process of rank aggregation which is interweave with the structure of skew-symmetric matrices. Recent development in the principles of matrix completion matrices is been applied and this idea gives rise to a new method for ranking a set of items. The core of this idea deals with the raking aggregation method which intimately describes a partially filled skew-symmetric matrix. The algorithm for matrix completion is raised to hold skew-symmetric data and use that to extract ranks for each item. This algorithm applies same strategy for both pairwise comparisons as well as for rating data. It becomes robust to noise and incomplete data as it is based on matrix completion.

A. Klementiev, D. Roth, and K. Small, describes the field of information retrieval, data mining, and natural language, many applications needs a ranking of instances which is not present in classification. Furthermore, a rank aggregation is a result of aggregating the results of the established ranking models into formalism and then result represents a novel unsupervised learning algorithm (ULARA) which gives a linear combination of individual ranking functions. These functions were developed based on the axiom of rewarding ordering agreement between the rankers.

A. Klementiev, D. Roth, and K. Small, produces a model which has to integrate the set of rankings often deals with aggregating and it only comes up when a certain ranked data is developed. Even though the various heuristic and supervised learning approaches to rank aggregation, a requirement of domain knowledge and supervised ranked data exists. Therefore, to solve this issue, a framework is proposed for learning aggregate rankings without supervision. This framework is instantiated for the cases of permutations and combinations of top-k lists.

III. ONLINE REPUTATION DEFENSE SYSTEM FOR FRAUD RANKING IN MOBILE APPS

In recent years, mobile app has been growing tremendously while boosting more than 400,000 applications like Apple app store and Google Android

market. This inclined growth of mobile apps makes it quite complicate to user for finding unique and trusted application in App stores. Thus to solve this important issue, existing marketing executives precisely use the App download history and ratings by the users to recommend the mobile applications which is totally trusted. Identifying ranking fraud is actually to identify ranking fraud of mobile apps within such leading sessions.

In this paper, an useful algorithm is used to discover the leading sessions based on the historical records and with the help of analysis of those records, it is proved that deceptive apps usually have different ranking patterns in each leading sessions as compared to the normal apps. Therefore, it is illustrated from those ranking records that some fraud is taking place in mobile app market and to restrict those frauds, three main evidences are developed to detect such fraud. As only ranking based evidences does not seems to be much sufficient to detect the fraud of mobile app, based on apps rating and review history some fraud evidences were discovered which showed anomaly patterns by those history. Specifically, an unsupervised evidence aggregation method is also proposed here to integrate all such types of evidences for evaluating the trustworthiness of leading sessions. The algorithm for detecting the mobile anomalies is structured in two phases namely, training and testing. The work flow of proposed approach is given in Fig.1.

Training Phase:

a) Pick the number of attributes n.
b) Name of the mobile user with their rating, the system is updated. Thus, every product with same username is trained.
c) And the same, IP address is also maintained and rated for every product.
d) For every data layer, apply feature selection techniques to the system.
e) Set the data in the data layer of trained models where the normal labelled data are passed to the next layer.

Testing Phase:

f) Test the arrived instances and name it as attack or normal.
g) If the sample is labelled as attack, then perform the action of blocking the attacker. If not, then move on to the upcoming layer.
h) Detaching the user from unfair ratings.

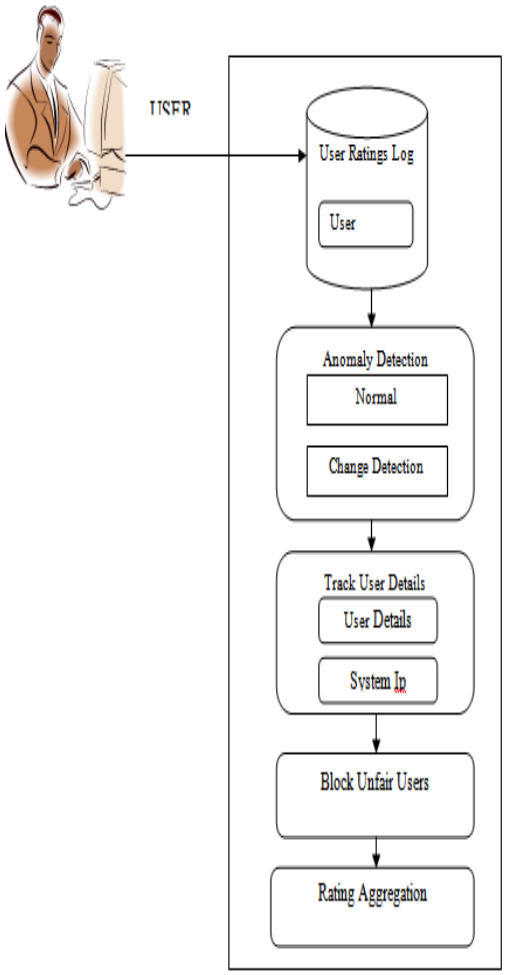

Fig.1. Proposed Workflow of discovering the fraud ranking

IV. EXPERIMENTAL RESULTS

In this section, we evaluate the performances of ranking fraud detection using real-world App data. The proposed approach is evaluated using Ranking based approaches and Rating based approaches.

In ranking based approaches, the positioning of the information are helpful for positioning extortion identification. It is not adequate to just utilize positioning based approaches. For the case, a few Apps made by the acclaimed designers, for example, Game space, might have a few driving occasions with vast estimations of the designers' validity and the "word-of-mouth" publicizing impact. In addition, a portion of the legitimate promoting

administrations, for example, "limited time discount", might likewise bring about critical ranking based confirmations. To comprehend this issue, we likewise interested to concentrate extortion confirmations from Apps' authentic rating records. In particular, after an App has been distributed, it can be evaluated by any client who downloaded it. In reality, client rating is a standout amongst the most critical components of Application promotion. An App which has higher rating might pull in more clients to download and can likewise be positioned higher in the pioneer board. Subsequently, appraising control is additionally an essential viewpoint of ranking extortion. Instinctively, if an App has ranking extortion in main sessions, the appraisals amid the time might have abnormality designs contrasted and its authentic evaluations, then which can be utilized for building rating base proofs.

In rating based approaches, other than ratings, the vast majority of the App stores likewise permit clients to compose the some printed remarks as App audits. Such audits can mirror the individual observations and use encounters of existing clients for specific portable Apps. Without a doubt, survey control is a standout amongst the most critical points of view of App positioning misrepresentation. Subsequently, fakers frequently post fake audits in the main sessions of a particular to swell the App download, and therefore drive the App's positioning the pioneer board.

V. CONCLUSION

In this paper, we developed a ranking fraud detection system for mobile Apps. Specifically, we first showed that ranking fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. Then, we identified ranking based evidences and rating based evidences for detecting ranking fraud. Moreover, we proposed fraud detection algorithm that integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps. A unique perspective of this approach is that all the evidences can be modeled by statistical hypothesis tests, thus it is easy to be extended with other evidences from domain knowledge to detect ranking fraud. Finally, we validate the proposed system with extensive experiments on real-world App data collected from the android markets. 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, 2011, pp. 181– 190.

[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, 2011, pp. 60–68.

[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 using 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. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 632–640.

[13] 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.

[14] G. Shafer, A Mathematical Theory of Evidence. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[15] Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen, "Discovery of Ranking Fraud for Mobile Apps", IEEE Transactions on Knowledge and Data Engineering, Vol. 27, No. 1, January 2015.