International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 14
October 2016

The Discovery Of The Order Of Fraud for Mobile Applications

[1]SHAIK AKBAR PASHA, [2]V RAMA RAO

[1] PG Scholar , Department Of CSE. Gandhi Academy Of Technical Education, Ramapuram (kattakommu Gudem), Chilkur(M), Kodad, Telangana 508206.

[2]Assistant Professor, Department Of CSE. Gandhi Academy Of Technical Education, Ramapuram (kattakommu Gudem), Chilkur(M), Kodad, Telangana 508206

ABSTRACT— 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.

1 INTRODUCTION

Structure of Data Mining Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

2 HOW DATA MINING WORKS?

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks.

3 EXTRACTING EVIDENCES FOR RANKING FRAUD DETECTION

In this section, we study how to extract and combine fraud evidences for ranking fraud detection.

3.1 RANKING BASED EVIDENCES

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 events for extracting fraud evidences.

By analyzing the Apps' historical ranking records, we observe that Apps' ranking behaviors in a leading event

always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase. Specifically, in each leading event, an App's ranking first increases to a peak position in the leaderboard (i.e., rising phase), then keeps such peak position for a period (i.e., maintaining phase), and finally decreases till the end of the event (i.e., recession phase). shows an example of different ranking phases of a leading event. Indeed, such a ranking pattern shows an important understanding of leading event. In the following, we formally define the three ranking phases of a leading event. Definition 3 (Ranking Phases of a Leading Event). Given a leading event e of App a with time

Note that, in Definition 3, DR is a ranking range to decide the beginning time and the end time of the maintaining phase. Teb and tec are the first and last time when the App is ranked into DR. It is because an App, even with ranking manipulation, cannot always maintain the same peak position (e.g., rank 1) in the leaderboard but only in a ranking range (e.g., top 25). If a leading session s of App a has ranking fraud, a's ranking behaviors in these three ranking phases of leading events in s should be different from those in a normal leading session. Actually, we find that each App with ranking manipulation always has an expected ranking target (e.g., top 25 in leaderboard for one week) and the hired marketing firms also charge money according to such ranking expectation (e.g., $1,000/day in top 25). Therefore, for both App developers and marketing firms, the earlier the ranking expectation meets, the more money can be

earned. Moreover, after reaching and maintaining the expected ranking for a required period, the manipulation will be stopped and the ranking of the malicious App will decrease dramatically. As a result, the suspicious leading events may contain very short rising and recession phases. Meanwhile, the cost of ranking manipulation with high ranking expectations is quite expensive due to the unclear ranking principles of App stores and the fierce competition between App developers. Therefore, the leading event of fraudulent Apps often has very short maintaining phase with high ranking positions.

4 DISCUSSION

Here, we provide some discussion about the proposed ranking fraud detection system for mobile Apps. First, the download information is an important signature for detecting ranking fraud, since ranking manipulation is to use so-called “bot farms” or “human water armies” to inflate the App downloads and ratings in a verym short time. However, the instant download information of each mobile App is often not available for analysis. In fact, Apple and Google do not provide accurate download information on any App. Furthermore, the App developers themselves are also reluctant to release their download information for various reasons. Therefore, in this paper, we mainly focus on extracting evidences from Apps’ historical ranking, rating and review records for ranking fraud

detection. However, our approach is scalable for integrating other evidences if available, such as the evidences based on the download information and App developers’ reputation.

5 EXPERIMENTAL RESULTS

In this section, we evaluate the performances of ranking fraud detection using real-world App data.

5.1 The Experimental Data

The experimental data sets were collected from the “Top Free 300” and “Top Paid 300” leaderboards of Apple’s App Store (U.S.) from February 2, 2010 to September 17, 2012.

The data sets contain the daily chart rankings1 of top 300 free Apps and top 300 paid Apps, respectively. Moreover, each data set also contains the user ratings and review information.

Apps with respect to different rankings in these data sets. In the figures, we can see that the number of Apps with low rankings is more than that of Apps with high rankings. Moreover, the competition between free Apps is more than that between paid Apps, especially in high rankings (e.g., top 25). Figs. 7a and 7b show the distribution of the number of Apps with respect to different number of ratings in these data sets. In the figures, we can see that the distribution of App ratings is not even, which indicates that only a small percentage of Apps are very popular.

5.2 Mining Leading Sessions

Here, we demonstrate the results of mining leading sessions in both data sets. Specifically, in Algorithm 1, we set the ranking threshold K_ ¼ 300 and threshold f ¼ 7. T

5.3 Human Judgement Based Evaluation

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 five human evaluators to validate the effectiveness of our approach Evidence Aggregation based Ranking Fraud Detection (EA-RFD). Particularly, we denote our approach with score based aggregation (i.e., Principle 1) as EA-RFD-1, and our approach with rank based aggregation (i.e., Principle 2) as EA-RFD-2, respectively.

6 PROPOSED SYSTEM

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.

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.

7 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, rating based evidences and review based evidences for detecting ranking fraud. Moreover, we proposed an optimization based aggregation method to integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps.

8 REFERENCE

[1] (2014). [Online]. Available: http://en.wikipedia.org/wiki/ cohen's_kappa

[2] (2014). [Online]. Available: http://en.wikipedia.org/wiki/ information_retrieval

[3] (2012). [Online]. Available: https://developer.apple.com/news/ index.php?id=02062012a

[4] (2012). [Online]. Available: http://venturebeat.com/2012/07/03/ apples-crackdown-on-app-ranking-manipulation/

[5] (2012). [Online]. Available: http://www.ibtimes.com/applethreatens-

crackdown-biggest-app-store-ranking-fraud-406764

[6] (2012). [Online]. Available: http://www.lextek.com/manuals/ onix/index.html

[7] (2012). [Online]. Available: http://www.ling.gu.se/lager/ mogul/porter-stemmer.

[8] 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.

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

[10] 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.

[11] 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.

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

[13] 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.

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

[15] J. Kivnen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," in Proc. 27th Annu. ACM Symp. Theory Comput., 1995, pp. 209–218.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

AUTHOR’S PROFILE:

SHAIK AKBAR PASHA

PG Scholar , Department Of CSE. Gandhi Academy Of Technical Education, Ramapuram (kattakommu Gudem), Chilkur(M), Kodad, Telangana 508206

V.RAMARAO

Assistant Professor, Department Of CSE. Gandhi Academy Of Technical Education, Ramapuram (kattakommu Gudem), Chilkur(M), Kodad, Telangana 508206