International Journal of Innovative Research in Engineering & Management (IJIREM)
ISSN: 2350-0557, Volume-3, Issue-3, May-2016

Discovery of Ranking of Fraud for Mobile Apps

Ranjitha.R
Computer Science and Engineering,
TRP Engineering College,
Trichy, India,

Mathumitha.K
Computer Science and Engineering,
TRP Engineering College,
Trichy, India,

Meena.S
Computer Science and Engineering,
TRP Engineering College,
Trichy, India,

S.Hariharan
Associate Professor,
Computer Science and Engineering,
TRP Engineering College,
Trichy, 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.

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

1. 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 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.

The rest of the paper is organized as follows: Section II, presents the literature survey over the related work. In section III, proposed system is presented in section IV, implementation for each modules. Finally, the section V concludes the review paper.

2. LITERATURE SURVEY
Leif Azzopardi et al. [2] studied an Investigating the Relationship between Language Model Perplexity and IR Precision Recall Measures the perplexity of the language model has a systematic relationship with the achievable precision recall performance though it is not statistically significant. A latent variable unigram based LM, which has been successful when applied to IR, is the so called probabilistic latent semantic indexing (PLSI).

Ee-Peng Lim et al. [12] 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.

David F. Gleich et al. [4] has done a survey on Rank Aggregation via Nuclear Norm Minimization the process of rank aggregation is

intimately intertwined with the structure of skew-symmetric matrices. To produces a new method for ranking a set of items. The essence of our idea is that a rank aggregation describes a partially filled skew-symmetric matrix. We extend an algorithm for matrix completion to handle skew-symmetric data and use that to extract ranks for each item.

Alexandre Klementiev, Dan Roth et al. [9] studied an Unsupervised Learning Algorithm for Rank Aggregation, (ULARA) which returns a linear combination of the individual ranking functions based on the principle of rewarding ordering agreement between the rankers.

3. PROPOSED SYSTEM

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 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. This challenge can be considered as detecting the local anomaly in place of global anomaly of mobile Apps. Second challenge, 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.

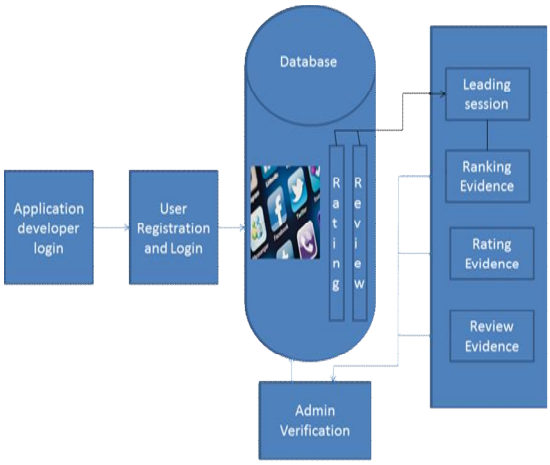

Figure 1.Architecture diagram

Mobile app stores launched many apps daily in the leader boards which shows the chart ranking of popular apps. The leader board is the important for promoting apps. Original application grade level decreases due to the arrival of fake apps. The users who are newly logging to the app stores, they decide based on the existing ranking, rating, reviews for the individual apps. 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 give 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, then the rating and reviews are correctly calculated.

4. IMPLEMENTATION

4.1 Identifying Leading Sessions

Ranking fraud usually happens in 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.

Mining Leading Sessions: 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.

4.2 Ranking Based Evidences

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 leader board (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).

4.3 Rating Based Evidences

The ranking based evidences are useful for ranking fraud detection. However, sometimes, it is not sufficient to only use ranking 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. Intuitively, if an App has ranking fraud in a leading session s, the ratings during the time period of s may have anomaly patterns compared with its historical ratings, which can be used for constructing rating based evidences.

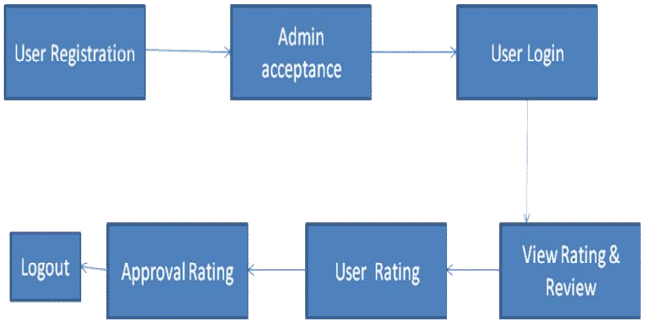

Figure 2.Rating based evidence

4.4 Review Based Evidences

Besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most important perspectives of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users often firstly 5, read its historical reviews to ease their decision making, and a mobile App contains more positive reviews may attract more users to download. Therefore, imposters often post fake reviews in the leading sessions of a specific App in order to inflate the App downloads, and thus propel the App's ranking position in the leader board. Although some previous works on review spam detection have been reported in recent years, the problem of detecting the local anomaly of reviews in the leading sessions and capturing them as evidences for ranking fraud detection are still under-explored.

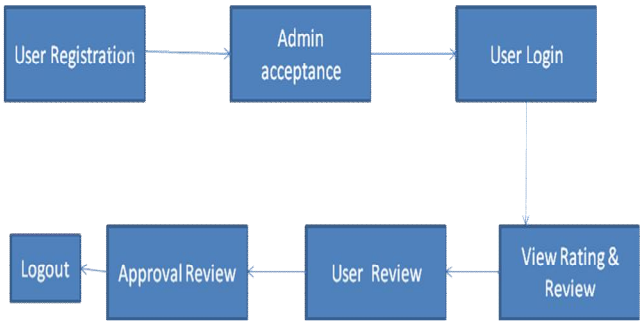

Figure 3.Rating based evidence

5. CONCLUSION AND FUTURE ENHANCEMENT

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 on admin verification method for evaluating the credibility of leading sessions from mobile Apps. An unique perspective of this approach is that all the evidences can be model by statistical hypothesis tests, thus it is easy to be extended with other evidences from domain knowledge to detect ranking fraud. The admin can detect the ranking fraud for mobile application. The Review or Rating or Ranking given by users is correctly calculated. Hence, a new user who wants to download an app for some purpose can get clear view about the available applications .Finally; we validate the proposed system with extensive experiments on real-world App data collected from the App store. Experimental results showed the effectiveness of the proposed approach.

In the future, we plan to study more effective fraud evidences and analyze the latent relationship among rating, review and rankings. Moreover, we will extend our ranking fraud detection approach with other mobile App related services, such as mobile Apps recommendation, for enhancing user experience.

REFERENCES

[1] Hengshu Zhu, Hui Xiong,Yong Ge, and Enhong Chen, "Discovery of Ranking Fraud for Mobile Apps"in Proc. IEEE 27th Int. Conf. Transactions on knowledge and data engineering, 2015, pp. 74-87.

[2] L. Azzopardi, M. Girolami, and K. V. Risjbergen, "Investigating the relationship between language model perplexity and in precision- recall measures," in Proc. 26th Int. Conf. Res. Develop. Inform. Retrieval, 2003, pp. 369–370.

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

[9] 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.

