Ranking Fraud Detection System for Mobile Apps

Akshay V. Metkar[1], Gaurav K. Shahane[2], Dipak N. Murtadak[3], Santosh D. Mutrak[4], Prof. R. B. Nangare[5]

[1,2,3,4] *BE Student, Computer Engineering Department, SVIT, Chincholi, Nashik, India.*
[5] *Professor, Computer Engineering Department, SVIT, Chincholi, Nashik, India.*

ABSTRACT

It became more and more frequent for App developers to use shameful means, such as inflating their Apps' sales or posting phony App ratings, to commit ranking fraud. In the mobile App market, ranking fraud refers to fraudulent or deceptive activities which have a purpose of bumping up the Apps in the popularity list. While the importance of preventing ranking fraud has been widely recognized, there is limited understanding and research in this area. In this paper, we provide a comprehensive way 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. 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. We also propose an optimization based aggregation method to integrate all the evidences for fraud detection. Finally, we evaluate the proposed system with real-world App data collected from the iOS App Store for a long time period. In the experiments, we validate the effectiveness of the proposed system, and show the scalability of the detection algorithm as well as some regularity of ranking fraud activities

Keyword: - *Mobile Apps, Ranking fraud detection, Evidence Aggregation, Historical ranking Records, rating and review.*

1. INTRODUCTION

Over the last few years the number of mobile Apps has been growing on a very large scale. At the end of April 2016 there is number of more than 3.4 million Applications at Apples App store and Google Play. Different App stores launched their leader board on daily basis to inspire the development of mobile Apps which displays the chart rankings of most popular Apps. In fact for promoting mobile Apps, leader board of apps is the most important ways in the market. An app ranking at the top on the leader board ultimately leads to a large number of downloads and million dollars in revenue. This results in exploring of different ways by the App developers like organizing promotional drives to advertise their Apps in order to get top position in App leader boards.

The very recent trend followed in market by the corrupt App developers for bumping up of an App is to use deceptive means to intentionally boost their apps. Lastly, the chart rankings on a App store are also manipulated. This is usually implemented by using so-called internet bots or human water armies to raise the App downloads, ratings and reviews in a very little time. Venture Beat is an article that reported, using ranking manipulation when an App was promoted, in Apples top free leader board it could be push forward from number 1,800 to the upmost 25 and new users more than 50,000-100,000 could be acquired within a couple of days.

In reality, such ranking fraud leads to great concerns to the industry of mobile App. For example, App developers who commit ranking fraud in the App store, Apple has warned of cracking down on them. As per the observation the mobile apps does not always ranked high in the leader boards, in fact in some leading events only. Collection of leading events of mobile Apps ultimately leads to different leading sessions. Thus, detecting ranking fraud of mob Apps happens in leading sessions and perhaps the process of detecting ranking fraud is done within the leading session of the mobile Apps. Especially, on the basis of historical ranking records of the mobile apps this paper proposes a simple and effective algorithm for the recognition of the leading sessions of each mobile App.

This is one of the evidence collected from historical ranking records of apps against fraud. Moreover, there are two more types of fraud evidences proposed on the basis of Apps rating and review history, which provides few anomaly patterns from Apps historical rating and review records. Additionally, system propose an unsupervised evidence-aggregation method to combine these three types of evidences collected for the assessment of credibility of leading sessions from mobile Apps. At the end, the proposed system is evaluated with app data collected from various resources.

2. LITERATURE SURVEY

M. N. Volkovs and R. S. Zemel [1] Many areas of study, such as information retrieval, collaborative filtering, and social choice face the preference aggregation problem, in which multiple preferences over objects must be combined into a consensus ranking. Preferences over items can be expressed in a variety of forms, which makes the aggregation problem difficult. In this work we formulate a flexible probabilistic model over pairwise comparisons that can accommodate all these forms. Inference in the model is very fast, making it applicable to problems with hundreds of thousands of preferences. Experiments on benchmark datasets demonstrate superior performance to existing methods. K. Shi and K. Ali, The Netflix competition of 2006 [2] has spurred significant activity in the commendations field, particularly in approaches using latent factor models [3,5,8,12] However, the near ubiquity of the Netflix and the similar MovieLens datasets1 may be narrowing the generality of lessons learned in this field. At GetJar, our goal is to make appealing recommendations of mobile applications (apps). For app usage, we observe a distribution that has higher kurtosis (heavier head and longer tail) than that for the aforementioned movie datasets. This happens primarily because of the large disparity in resources available to app developers and the low cost of app publication relative to movies.

In this paper we compare a latent factor (PureSVD) and a memory-based model with our novel PCA-based model, which we call Eigenapp. We use both accuracy and variety as evaluation metrics. PureSVD did not perform well due to its reliance on explicit feedback such as ratings, which we do not have. Memory-based approaches that perform vector operations in the original high dimensional space over-predict popular apps because they fail to capture the neighborhood of less popular apps. They have high accuracy due to the concentration of mass in the head, but did poorly in terms of variety of apps exposed. Eigenapp, which exploits neighborhood information in low dimensional spaces, did well both on precision and variety, underscoring the importance of dimensionality reduction to form quality neighborhoods in high kurtosis distributions.

A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly [3], In this paper, we continue our investigations of "web spam": the injection of artificially-created pages into the web in order to influence the results from search engines, to drive traffic to certain pages for fun or profit. This paper considers some previously-undescribed techniques for automatically detecting spam pages, examines the effectiveness of these techniques in isolation and when aggregated using classification algorithms. When combined, our heuristics correctly identify 2,037 (86.2%) of the 2,364 spam pages (13.8%) in our judged collection of 17,168 pages, while misidentifying 526 spam and non-spam pages (3.1%).

3. SYSTEM ARCHITECTURE

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.

In Ranking Based 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.

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. 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. 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.

The proposed framework is scalable and can be extended with other domain generated evidences for ranking fraud detection.

Experimental results show the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity of ranking fraud activities.

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).

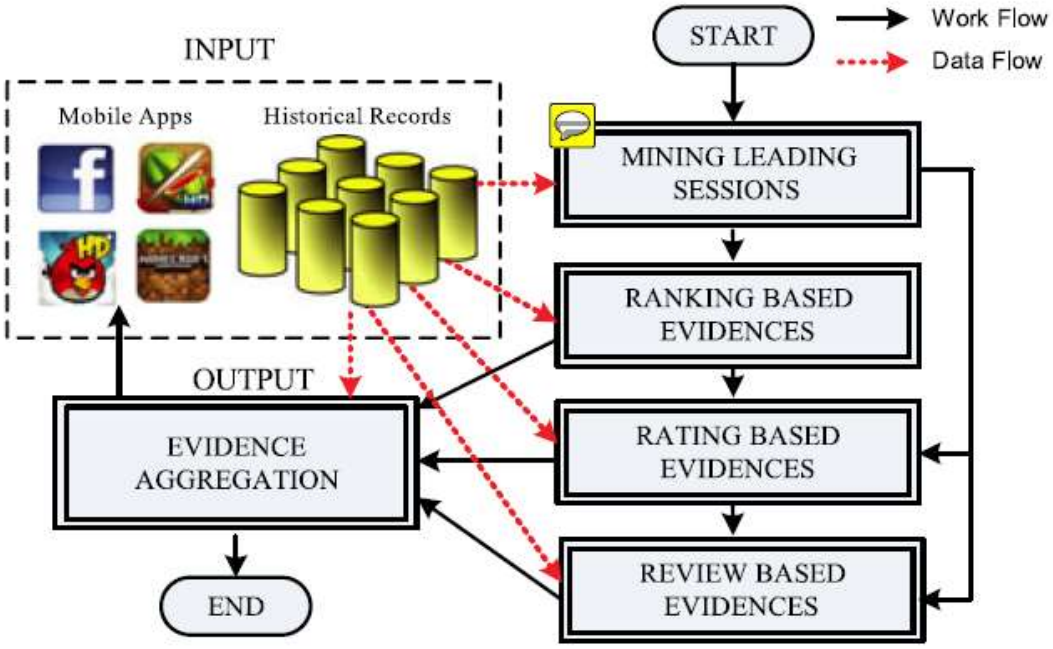

Figure 1: System Architecture.

Modules:

- Mining Leading Sessions
- Ranking Based Evidences
- Rating Based Evidences
- Review Based Evidences
- Evidence Aggregation

Modules Description:

Mining Leading Sessions

In the first module, we develop our system environment with the details of App like an app store. Intuitively, the leading sessions of a mobile App represent its periods of popularity, so the ranking manipulation will only take place in these leading sessions. Therefore, the problem of detecting ranking fraud is to detect fraudulent leading sessions. Along this line, the first task is how to mine the leading sessions of a mobile App from its historical ranking records. 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.

Ranking Based Evidences

In this module, we develop Ranking based Evidences system. By analyzing the Apps' historical ranking records, web serve 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).

Rating Based Evidences

In the third module, we enhance the system with Rating based evidences module. The ranking based evidences are useful for ranking fraud detection. However, sometimes, it is not sufficient to only use ranking based evidences. For example, some Apps created by the famous developers, such as Gameloft, may have some leading events with large values of u1 due to the developers' credibility and the "word-of-mouth" advertising effect. Moreover, some of the legal marketing services, such as "limited-time discount", may also result in significant ranking based evidences. To solve this issue, we also study how to extract fraud evidences from Apps' historical rating records.

Review Based Evidences

In this module, we add the Review based Evidences module in our system. 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 first 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.

Evidence Aggregation

In this module, we develop the Evidence Aggregation module to our system. After extracting three types of fraud evidences, the next challenge is how to combine them for ranking fraud detection. Indeed, there are many ranking and evidence aggregation methods in the literature, such as permutation based models score based models and Dumpster-Shafer rules. However, some of these methods focus on learning a global ranking for all candidates. This is not proper for detecting ranking fraud for new Apps. Other methods are based on supervised learning techniques, which depend on the labeled training data and are hard to be exploited. Instead, we propose an unsupervised approach based on fraud similarity to combine these evidences.

4. CONCLUSIONS

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. An 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 Apple's 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.

5. ACKNOWLEDGEMENT

We take this opportunity to express our hearty thanks to all those who helped us in the completion of the paper. We express our deep sense of gratitude to our guide Prof. R. B. Nangare, Asst. Prof., Computer Engineering Department, Sir Visvesvaraya Institute of Technology, Chincholi for his guidance and continuous motivation. We gratefully acknowledge the help provided by him on many occasions, for improvement of this project report with great interest. We would be failing in our duties, if we do not express our deep sense of gratitude to Prof. S. M. Rokade, Head, Computer Engineering Department for permitting us to avail the facility and constant encouragement. Lastly we would like to thank all the staff members, colleagues, and all our friends for their help and support from time to time.

6. REFERENCES

[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.