ISSN- 2394-5125 VOL 02, ISSUE 03, 2015

Mobile Apps Ranking Fraud Discovery based on comparison threshold

A. Ramaswami Reddy

Assistant Professor, Computer Science Engineering, Vignan's Foundation for Science, Technology & Research (Deemed to be University) Deemed university in Guntur, Andhra Pradesh

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. 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 addition, we 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.

Keywords: - Ranking, Rating, App.

1. INTRODUCTION

The number of mobile Apps has grown at a breathtaking rate over the past few years. For example, as of the end of April 2013, there are more than 1.6 million Apps at Apple's App store and Google Play. To stimulate the development of mobile Apps, many App stores launched daily App leaderboards, which demonstrate the chart rankings of most popular Apps. Indeed, the App leaderboard is one of the most important ways for promoting mobile Apps. A higher rank on the leaderboard usually leads to a huge number of downloads and million dollars in 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 leaderboards. However, as a recent trend, instead of relying on traditional marketing solutions, shady App developers resort to some fraudulent means to deliberately boost their Apps and eventually manipulate the chart rankings on an App store. This is usually implemented by using so-called "bot farms" or "human water armies" to inflate the App downloads, ratings and reviews in a very short time. For example, an article from Venture Beat reported that, when an App was promoted with the help of ranking manipulation, it could be propelled from number 1,800 to the top 25 in Apple's top free leaderboard and more than 50,000-100,000 new users could be acquired within a couple of days. In fact, such ranking fraud raises great concerns to the mobile App industry. For example, Apple has warned of cracking down on App developers who commit ranking fraud in the Apple's App store. In the literature, while there is some related work, such as web ranking spam detection online review spam detection and mobile App recommendation the problem of detecting ranking fraud for mobile Apps is still under-explored. To fill this crucial void, in this paper, we propose to develop a ranking fraud detection system for mobile Apps. Along this line, we identify several important challenges. First, ranking fraud does not always happen in the whole life cycle of an App, so we need to detect the time when fraud happens. Such challenge can be regarded as detecting the local anomaly instead of global anomaly of mobile Apps. Second, due to the huge number of mobile

ISSN- 2394-5125 VOL 02, ISSUE 03, 2015

Apps, it is difficult to manually label ranking fraud for each App, so it is important to have a scalable way to automatically detect ranking fraud without using any benchmark information. Finally, due to the dynamic nature of chart rankings, it is not easy to identify and confirm the evidences linked to ranking fraud, which motivates us to discover some implicit fraud patterns of mobile Apps as evidences. Indeed, our careful observation reveals that mobile Apps are not always ranked high in the leaderboard, but only in some leading events, which form different leading sessions. Note that we will introduce both leading events 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 characterize 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 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.

2. RELATED WORK

In the literature, while there are some related work, such as web ranking spam detection, online review spam detection and mobile App recommendation, the problem of detecting ranking fraud for mobile Apps is still under-explored. Generally speaking, the related works of this study can be grouped into three categories. The first category is about web ranking spam detection. The second category is focused on detecting online review spam. Finally, the third category includes the studies on mobile App recommendation although some of the existing approaches can be used for anomaly detection from historical rating and review records, they are not able to extract fraud evidences for a given time period (i.e., leading session). Cannot able to detect ranking fraud happened in Apps' historical leading sessions. There is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud. 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

ISSN- 2394-5125 VOL 02, ISSUE 03, 2015

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).

3. IMPLEMENTATION

Admin:

Admin is main user of an application. Admin can handle and accept or reject apps which uploaded by providers. Admin performs ranking based evidences, rating based evidences and review based evidences, by modeling Apps' ranking. Admin find the App Providers who commit ranking fraud.

App Providers

App Providers are a data providers of this application, they uploads their developed apps in app store for promoting. App Providers 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.

User:

User is an end user of our application, after completion of registration he can login to access the application data. He can download the apps. App 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.

Proposed Algorithm:

Input: N = 1, 2, 3, 4….n (number of apk's)

Input: R ∈ r where r =1, 2,3,4,5.

Input: ϕ - is the comparison threshold

Initialization Steps:

Step1: $\sum_{i=1}^{n}$ where n = no of Apk's. count

Step2: r ∈ N giving rating

Step 3: **for each** N ∈ r=[1,|R] do,

if R ⊂ r = ϕ here compares user's and local apk's ranking and downloads. Will download → Ranking depends on rating THRESHOLD

Else if R ⊆ r = ϕ // both user's and local apk's ranking are equal. NOT FRAUD

Else

Download N=1, 2, 3, 4, 5,…n and rate the Apk's.

4. EXPERIMENTAL RESULTS

Fig:-1 App list.

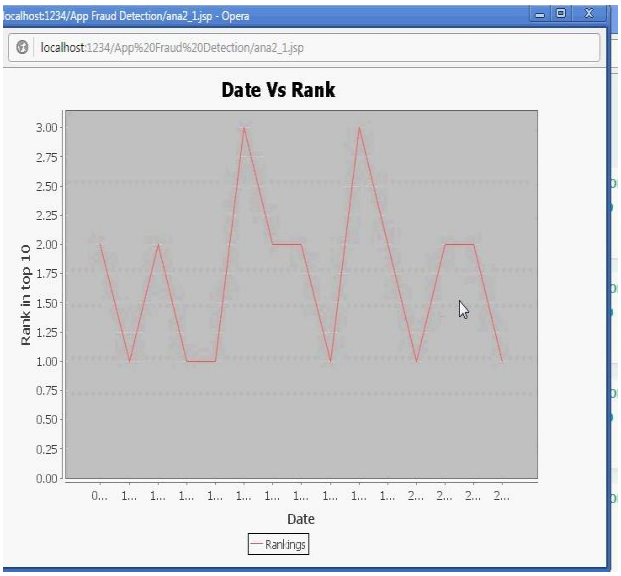

Fig:-2 Date Rank-1

ISSN- 2394-5125 VOL 02, ISSUE 03, 2015

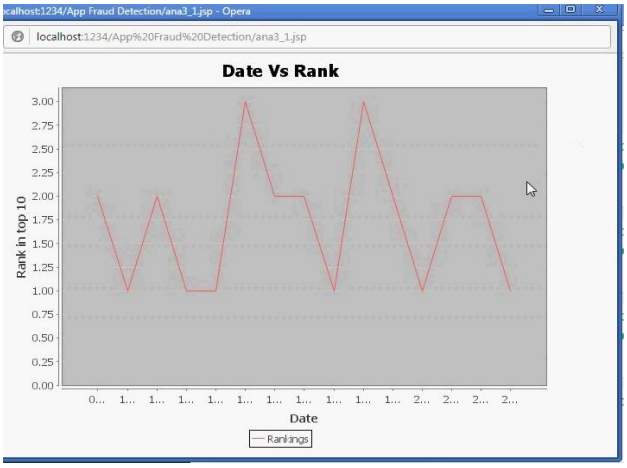

Fig:-3 Date Rank-2

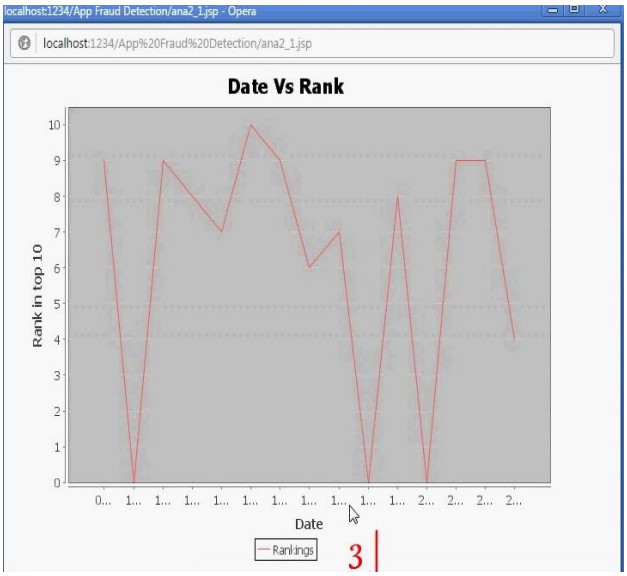

Fig:-3 Date Rank-3

5. 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. 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 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 raid detection approach with other mobile App related services, such as mobile Apps recommendation, for enhancing user experience.

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-fra ud-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.