International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307

Detection of Fraud in the ranking of Mobile Apps

Nalluri Sunny*, Kodali Eswar, Mastan MD Meera Durga
Computer Science and Engineering Department, V R Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India

ABSTRACT

The ranking fraud in the mobile Apps market leads to misleading and illusive activities which is used in keeping more and more Apps in the popularity list. The developers such as posting unauthentic App ratings, for the purpose of ranking, used many of the misguiding means. In this paper a comprehensive view of ranking fraud and a detection system for the mobile apps is given. The proposed method is based on the leading sessions of the mobile Apps and locating the ranking fraud by the active periods in them. Those leading sessions were used for detecting the global and local App rankings. Evidences like ranking based evidences, rating based evidences and review based evidences through statistical hypotheses tests were used. Finally, we evaluated the proposed system with real-world App data collected from the iOS App Store for a long time period. In the experiments, we also validated the effectiveness of the proposed system, and the scalability of the detection algorithm as well as some regularity of ranking fraud activities.

Keywords: Rate based, Ranking based evidences, Mining leading sessions

I. INTRODUCTION

For the promotion of mobile Apps, certain app stores have launched daily App leader boards, to show the chart rankings of their popular Apps as it is the best way for promoting the Apps. So a higher rank on the leader boards leads to large number of downloads which draws a revenue of million dollars. To promote their Apps, App developers are exploring different ways like advertising campaigns etc., to have their Apps ranked in high ratings. Due to this competition some shady app developers instead of relying on traditional marketing companies, they make some fradulent means to boost up their Apps which leads to the manipulation of chart rankings on the App store. For example, an article from VentureBeat [1] stated 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 leader board and more than 50,000- 100,000 new users could be acquired within a couple of days. 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 underexplored. To fill this crucial void, in our project, we have proposed to develop a ranking fraud detection system for mobile Apps.

The organization of this document is as follows. In Section II describes about basic concepts, section III describes about literature survey, section IV describes about methodology, section V describes about results, section VI describes about conclusion and section VII describes about references.

II. Basic concepts

The main objective is to develop a system for detecting the fraud behind ranking the various mobile apps available.

1) 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.
2) Second, due to the huge number of mobile 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.
3) Finally, some implicit fraud patterns of mobile Apps as evidences were discovered.

III. Literature Survey

Web ranking spam refers to any serious actions, which bring to selected Web pages an unjustifiable and favorable importance. In this, the problem of unsupervised web spam detection is studied. They introduce the concept of spamicity to measure how likely a page is a spam. Spamicity is more flexible and user controllable measure than the traditional supervised classification methods. They propose efficient online link spam and term spam detection methods using spamicity. This method does not need training and is also cost effective. A real data set is used to evaluate the effectiveness and the efficiency.
Ntoulas et al. [2] have studied various aspects of content-based spam on the Web and presented a number of heuristic methods for detecting content-based spam. In this paper, they continue 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, which examines the effectiveness of these techniques in isolation with classification algorithms.
Recently, Spirin et al. [3] have reported a survey on Web spam detection, which comprehensively introduces the principles and algorithms in the literature. Indeed, the work of Web ranking spam detection is mainly based on the analysis of ranking principles of search engines, such as PageRank and query term frequency. This is different from ranking fraud detection for mobile Apps. They categorized all existing algorithms into three categories based on the type of information they use: content based methods, link-based methods, and methods based on non-traditional data such as user behavior, clicks, and HTTP sessions. In turn, there is a sub categorization of link based category into five groups based on ideas and principles used: labels propagation, link pruning and reweighting, labels refinement, graph regularization, and feature based. Here effectiveness of the system is less, wastage of the used resources is high, and information may become unreliable, if proper update is not done.
Lim et al. [4] have identified several representative behaviors of review spammers and model these behaviors to detect the spammers. This paper aims to detect users generating spam reviews or review spammers. They identify several characteristic behaviors of review spammers and model these behaviors so as to detect the spammers. In particular, authors seek to model the following behaviors. First, spammers may target specific products or product groups in order to maximize their impact. Second, they tend to deviate from the other reviewers in their ratings of products. They propose scoring methods to measure the degree of spam for each reviewer and apply them on an Amazon review dataset. Authors then select a subset of highly suspicious reviewers for further scrutiny by user evaluators with the help of a web based spammer evaluation software specially developed for user evaluation experiments.

IV. METHODOLOGY

Mining Leading Sessions

In the first module, system environment with the details of App like an app store is developed. The

leading sessions of a mobile App represent its periods of popularity, so the ranking manipulation will only take place in these leading sessions[5]. 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.

Leading Event

Leading Event of an app is that which contains the rankings of the app in a certain time range, which satisfies some conditions.

Leading Session

Leading Session of an app contains a time range and adjacent leading events. The leading sessions of a mobile app represent its periods of popularity [6], so the ranking manipulation will only take place in these leading sessions. Therefore, the problem of detecting ranking fraud is to detect fraudulent leading session. Here we should first analyze the basic characteristics of leading events for extracting fraud evidences. App's ranking behavior in a leading event always satisfies a specific ranking pattern. Usually apps have three different ranking phases, namely

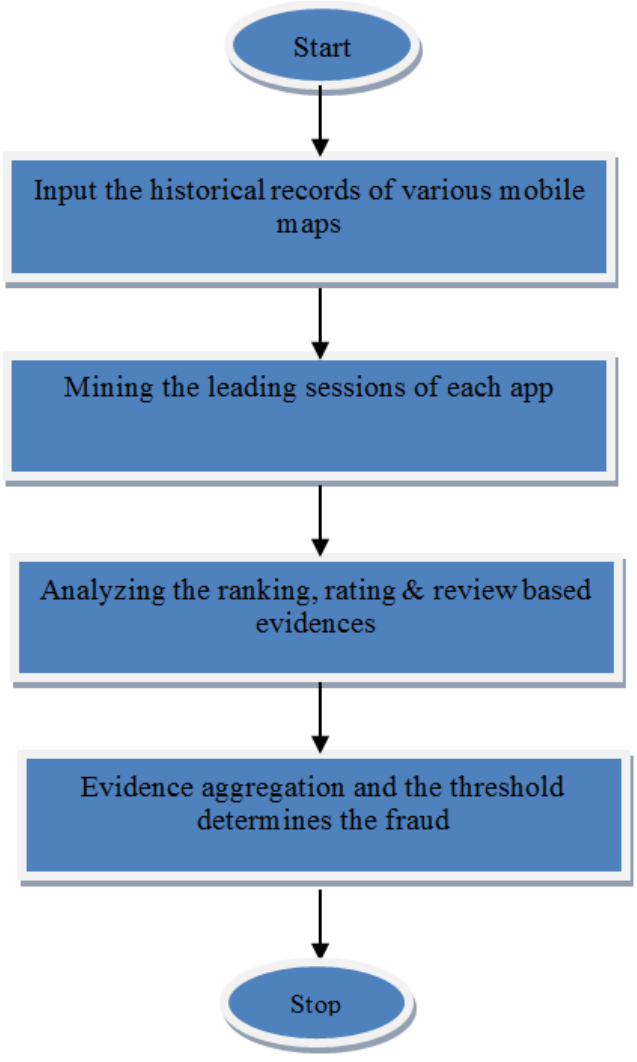
Rising phase: Increasing to a peak position

Maintenance phase: Maintains the peak position for a period

Recession phase: Decreasing to the end if a leading session of App has ranking fraud, app's ranking behavior in these three ranking phases of leading events in leading sessions should be different from those in a normal leading session.

Ranking Based Evidences

In this module, Ranking based Evidences system is developed. 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[7]. 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).

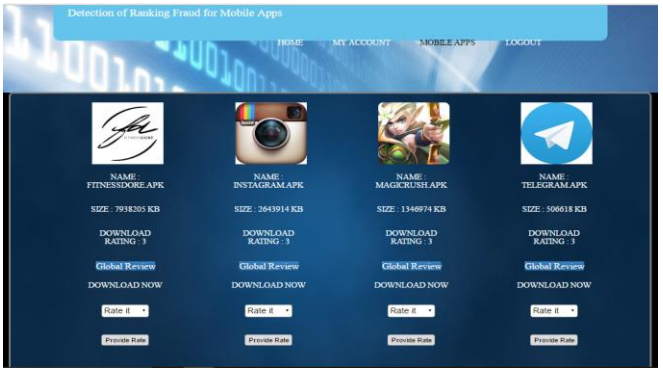


Rating Based Evidences

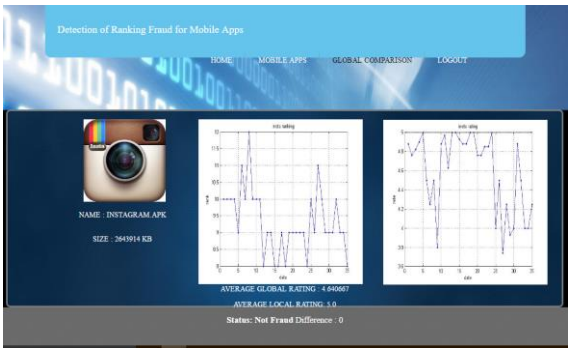
In the third module, the system with Rating based evidences module is enhanced. The ranking based evidences are useful for ranking fraud detection.

However, sometimes, it is not sufficient to only use ranking based evidences [8]. 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 [9]. To solve this issue, we also study how to extract fraud evidences from Apps' historical rating records.

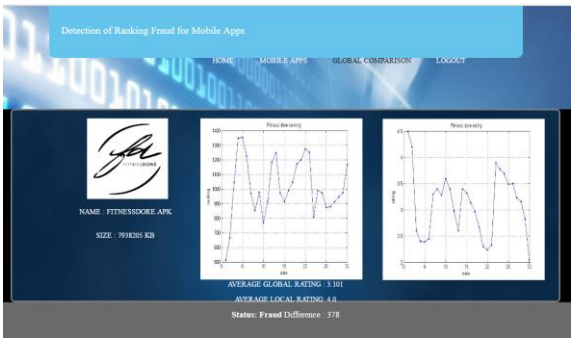
V. RESULTS



One of the Non-fraud App



One of the fraud App:



The above results show that based on the ranking and rating based differences the average global and local ratings were calculated and the status shows that there is fraud in the app.

VI. CONCLUSION

Along with the increase in mobile apps since few years, the ranking fraud for these apps also gradually increased. Due to the ranking fraud in mobile apps, the app ranking can change from few thousands to few hundreds. As the users generally download an app by seeing its rank, the companies, which are developing apps, are gaining millions of rupees illegitimately [10]. So, to control this 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 .We showed the fraud in the form of graphs from which we can draw conclusions. In order to provide this information to users, we created a website through which the users

can register, login and know the fraud in apps. The frequent users need not register, they can directly login. If the user is interested in any app, they can download the app. Due to this, any user before installing an app, can know whether the rank given to the app is correct or not. The user will not get manipulated by the fraud ranking.

VII. FUTURE WORK

Business and research communities are significantly attracted by this fake review detection. As increased popularity, mobiles are major target for malicious applications. Main challenge is to detect and remove malicious apps from mobile app market[11]. Thus, there is need to have novel system to effectively analyze fraud apps. Future the system can also be extended to make a detection system that tracks the online campaigning on social media. Online campaigning may be used to gain a particular benefit that may be of business, politics or something else[12]. For future work, this technique can be implemented on some more apps. Increasing the number of apps used for the process, can improve the accuracy. This can also be extendsed to take data online and display the graph to the users automatically

VIII. REFERENCES

[1]. L Azzopardi, M Girolami, et al. Investigating the relationship between language model perplexity and IR precision-recall measures, in Proc. 26th Int. Conf. Res. Develop. Inform. Retrieval, 2016; 369–370.

[2]. Rahman, S Huang, HV.Faloutsos. Detecting malicious Facebook applications. IEEE transactions on networking volume, 2015.

[3]. Z Hengshu, X Hui Xiong, et al. Discovery of ranking fraud for mobile apps. IEEE Transactions on knowledge and data engineering, 2014.

[4]. Exploiting enriched contextual information for mobile app classification, H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian. In Proceedings of the 21st ACM international conference on Information and knowledge management, CIKM ’12, pages 1617–1621, 2015.

[5]. N. Spirin and J. Han, "Survey on web spam detection: Principles and algorithms," SIGKDD Explor. Newslett., vol. 13, no. 2, pp. 50–64, May 2016.

[6]. E.-P. Lim,V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers use rating behaviors," in Proc.19thACMInt. Conf. Inform. Knowl. Manage., 2014, pp. 939–948.

[7]. K. Shi and K. Ali. Getjar mobile application recommendations with very sparse datasets. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ’12, pages 204–212, 2016.

[8]. A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in Proc. 15th Int. Conf. World Wide Web, 2014, pp. 83–92

[9]. B. Yan and G. Chen. Appjoy: personalized mobile application discovery. In Proceedings of the 9th international conference on Mobile systems, applications, and services, MobiSys ’11, pages 113–126, 2015.

[10]. B. Zhou, J. Pei, and Z. Tang, "A spamicity a pproach to web spam detection," in Proc. SIAM Int. Conf. Data Mining, 2016. 66

[11]. A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh. Spotting opinion spammers using behavioral footprints. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ’13, 2016.

[12]. H. Zhu, H. Xiong, Y. Ge, and E. Chen. Ranking fraud detection for mobile apps: A holistic view. In Proceedings of the 22nd ACM international conference on Information and knowledge management, CIKM ’13, 2016.