International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Survey: Fraud Detection and Discovery of Mobile Apps

Urmila Aware[1], D. O. Shamkuwar[2]

[1]Student, Dept of Computer Engineering, Flora Institute of Technology Pune, Maharashtra, India
[2]Assistant Professor, Dept of Computer Engineering, Flora Institute of Technology Pune, Maharashtra, India

Abstract: ***Now a days everyone is using smart phones. There is need of various applications to be installed on smart phone. To download application smart phone user has to visit play store such as Google Play Store. Mobile App is a very popular and well known concept due to the rapid advertisement in the mobile technology and mobile devices. When user go to the play store then he is see the various application lists. This list is built on the basis of advertisement. Usually user doesn't have enough information about the application (i.e. which applications are useful or not). Then a user sees that list and downloads the applications. But sometimes that the downloaded application not useful. That means it is fraud in mobile application list. To avoid this fraud, we provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile Apps. We first propose to accurately locate the ranking fraud by mining the active periods, namely leading sessions, of mobile Apps. Then we investigate five types of evidences, i.e., ranking based evidences, rating based evidences, review based evidences, recommendation based fraud detection, enhancing the algorithm of fraud detection, by modeling Apps' ranking, rating and review behaviors through statistical hypotheses tests. Using these five evidences finally we are calculating aggregation. We evaluate our application with real world data collected form play store for long time period.***

Keywords: Mobile Apps, Ranking Fraud Detection, Evidence Aggregation, Historical Ranking Records, Rating and Review, Recommendation app, Ad History etc.

1. Introduction

Over the past few years the number of mobile Apps has grown at a breathtaking rate. For example, there are more than 1.6 million Apps at Google Play and Apple's App store, as of the end of July 2015. To stimulate the development of mobile Apps, many App stores launched daily App leader boards, which demonstrate the chart rankings of most popular Apps. The recent trend in market used by the dishonest App developers for App boosting is to use fraudulent means to consciously boost their apps. At last, they also distort the chart rankings on a App store. 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. For example, Venture Beat [1] reported that, when an App was promoted using ranking manipulation, it could be precipitated from number 1,800 to the upmost 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. The quantity of mobile Apps has developed at an amazing rate in the course of recent years. For instances, the growth of apps were increased by 1.6 million at Apple's App store and Google Play. To increase the development of mobile Apps, many App stores launched daily App leader boards, which demonstrate the chart rankings of most popular Apps and adds more ads in their app to earn. Indeed, the App leader board and mobile ads are one of the most important ways for promoting mobile Apps and for earning respectively. A higher rank on the leader board usually leads to a huge number of downloads and million dollars in revenue with mobile ads. 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 leader boards and likewise can earn more by showing multiple ads. 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. Therefore, the main target is to detect ranking fraud of mobile Apps within leading sessions and fining fake ads. First propose an effective algorithm to identify the leading sessions of each App based on its historical ranking records. Then, with the analysis of Apps' ranking behaviors, find out the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, some fraud evidences are characterized from Apps' historical ranking records. Then three functions are developed to extract such ranking based fraud evidences. Therefore, further four types of fraud evidences are proposed based on Apps' rating and review history, recommendation and ad fraud history which reflect some anomaly patterns from Apps' historical rating and review records. In addition, to integrate these types of evidences, an unsupervised evidence-aggregation method is developed which is used for evaluating the credibility of leading sessions from mobile Apps.

2. Literature Survey

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 categorize 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, 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.[2] Lim et al. [4] have identified several representative behaviors of review spammers and model these behaviors to detect the

Licensed Under Creative Commons Attribution CC BY

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. In the literature, while there are some related work, such as web ranking spam detection [2], [5], [6], online review spam detection [7], [8], [9], and mobile App recommendation the problem of detecting ranking fraud for mobile Apps is still under-explored.

3. Proposed System

We can say that, ranking fraud usually happens in these leading sessions. So, detecting ranking fraud of mobile Apps is truly to detect ranking fraud within leading sessions of mobile Apps. Specifically, we first propose a simple yet convincing algorithm to identify the leading sessions of each App based on its historical ranking records. At that time, with the analysis of Apps' ranking behaviours, we find that the fraudulent Apps often have different ranking patterns in leading session liken with normal Apps. Thus, we characterize some evidences which is fraud from App's historical ranking records, and develop three task to extract such ranking based fraud evidences. Nonetheless, the evidences of ranking based can be affected by App developers' reputation and some legitimate marketing campaigns, such as "limited period discount and more". As a conclusion, it is not sufficient to only use ranking based evidences. Therefore, we further suggest two types of fraud evidences based on App's review and rating history, which reflect some discrepancy patterns from Apps' historical rating and review records.

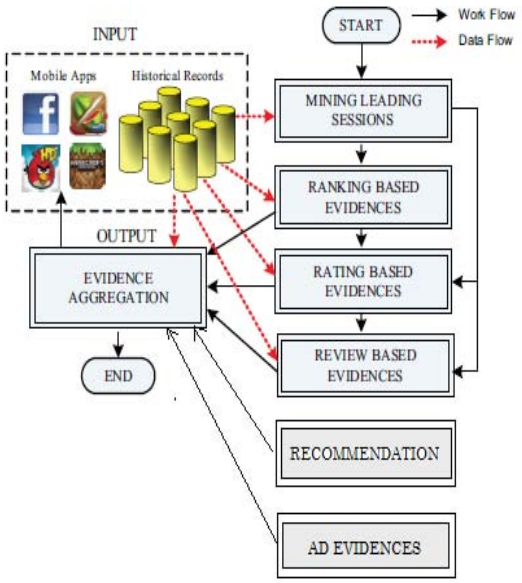


Figure 1: The framework of our ranking fraud detection system for mobile Apps.

There are two main phases for detecting the ranking fraud:

1) Identifying the leading sessions for mobile apps
2) Identifying evidences for ranking fraud detection

i) Identifying Leading Sessions for Mobile Apps

The App leaderboard of play storemanifest top K popular Apps with respect to various categories, such as "Top Paid Apps" and "Top Free Apps". Besides, the leaderboard is usually updated periodically (e.g., daily). So, each individual mobile App *a* has many historical ranking records which can be denoted as a time series, $Ra = \{ra1, \ldots, rai, \ldots, ran\}$, where $r\,i \in \{1, \ldots, K, +\infty\}$. the ranking of a at time stamp *ti*; $+\infty$ means *a* is not ranked in the top *K* list; *n* indicates the number of all ranking records. the leading sessions of a mobile App represent its times of popularity, therefore the ranking manipulation will only take place in these leading sessions. So, the issue of detecting ranking fraud is to detect fraudulent leading sessions. Along this line, the first job is how to mine the leading sessions of a mobile App from its historical ranking records.

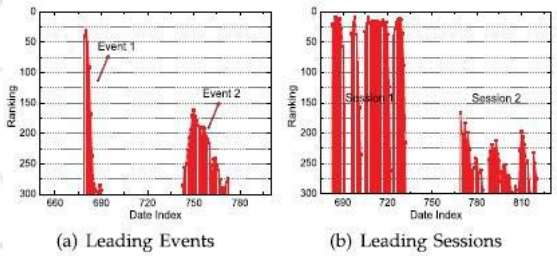


Figure 2: (a) Example of leading events; (b) Example of leading sessions

ii)) Identifying evidences for ranking fraud detection

Let us see these in brief:

1) Ranking based evidences:

It concludes that leading session comprises of various leading events. Hence by analysis of basic behaviour of leading events for finding fraud evidences and also for the app historical ranking records.

2) Rating based evidences:

For ranking fraud detection are uses the ranking based evidences. However, sometimes, it is not sufficient to only use ranking based evidences. As we know that rating is been done after downloading it by the user, and if the rating is high in leaderboard considerably that is attracted by most of the mobile app users.

3) Review based evidences:

Review manipulation is one of the most valuable perspective of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users usually first read its historical reviews to ease their decision making, and a mobile App contains more encouraging reviews may captivate more users to download.

4) Evidence Aggregation:

After extracting three types of fraud evidences, the next challenge is how to combine them for ranking fraud detection.

In addition, there are many methods of ranking and evidence aggregation in the literature, such as permutation based

Licensed Under Creative Commons Attribution CC BY

models [10], [11], score based models and Dempster Shafer rules [12].

5) Recommendation Based Fraud:
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 and **for recommendation** of 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.

4. Experimental Result

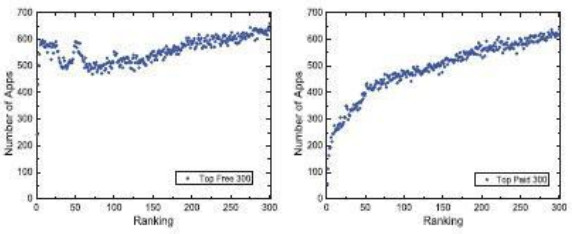
In this section, we evaluate the performances of ranking fraud detection using real-world App data.

A. 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. Furthermore, each data set also contains the user ratings and review information. Figs. 5a and 5b indicates the distributions of the number of Apps with respect to different rankings in these data sets. In these figures, we can notice that the number of Apps with low rankings is more than that of Apps with high rankings. Additionally, the competition between free Apps is more than that between paid Apps, especially in high rankings (e.g., top 25). Figs. 6a and 6b show the distribution of the number of Apps with respect to different number of ratings in these data sets. In these figures, we can notice that the distribution of App ratings is not even, which shows that only a small percentage of Apps are very popular.

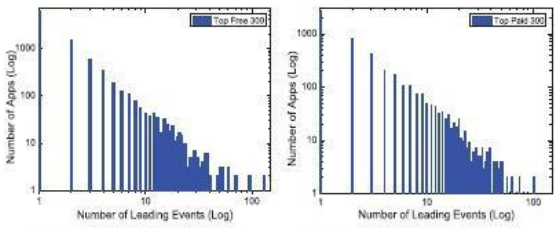
B. Mining Leading Sessions

Here, we indicate the results of mining leading sessions in both data sets. Specifically, in Algorithm 1, we set the ranking threshold $k^* = 300$ and threshold $\phi = 7$. This denotes two adjacent leading events can be segmented into the same leading session if they happen within one week of each other. Figs. 7 and 8 show the distributions of the number of Apps with respect to different numbers of contained leading events and leading sessions in both data sets. In these figures, we can notice that only a few Apps have many leading events and leading sessions. The average numbers of leading events and leading sessions are 2:69 and 1:57 for free Apps, and 4:20 and 1:86 for paid Apps. Moreover, Figs.9a and 9b show the distribution of the number of leading sessions with respect to different numbers of contained leading events in both data sets. In the figures, we can find only a few leading sessions contain many leading events.



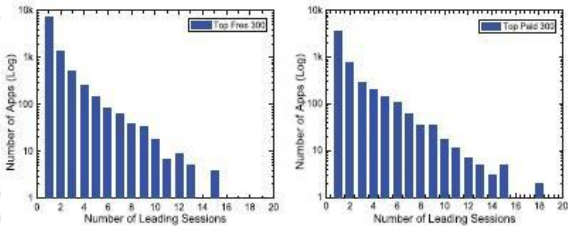
(a) Top Free 300 data set (b) Top Paid 300 data set

Figure 5: The distribution of the number of Apps w.r.t different rankings.



(a) Top Free 300 data set (b) Top Paid 300 data set

Figure 7: The distribution of the number of Apps w.r.t different numbers Of ratings.



(a) Top Free 300 data set (b) Top Paid 300 data set

Figure 8: The distribution of the number of Apps w.r.t different number of leading sessions.

5. Conclusion

We conclude that, to develop a ranking fraud detection system for mobile Apps. we first discover that ranking fraud occur in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. By mining the leading sessions of mobile Apps, we aim to locate the ranking fraud. 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.

References

[1] (2012). [Online]. Available: http://venturebeat.com/2012/07/03/ apples-crackdown-on-app-ranking-manipulation/
[2] N. Spirin and J. Han. Survey on web spam detection: principles and algorithms. SIGKDD Explor. Newsl., 13(2):50–64, May 2012.
[3] https://developer.apple.com/news/index.php?id=02062012a.
[4] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting product review spammers using rating behaviors. In Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM '10, pages 939–948, 2010.

Licensed Under Creative Commons Attribution CC BY

[5] M. N. Volkovs and R. S. Zemel. A flexible generative model for preference aggregation. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 479–488, 2012.
[6] B. Zhou, J. Pei, and Z. Tang. A spamicity approach to web spam detection. In *Proceedings of the 2008 SIAM International Conference on Data Mining*, SDM'08, pages 277–288, 2008.
[7] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 939–948, 2010.
[8] Z. Wu, J. Wu, J. Cao, and D. Tao. Hysad: a semi-supervised hybrid shilling attack detector for trustworthy product recommendation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pages 985–993, 2012.
[9] S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pages 823–831, 2012.
[10] B. Yan and G. Chen, "AppJoy: Personalized mobile application discovery," in Proc. 9th Int. Conf. Mobile Syst., Appl., Serv., 2011, pp. 113–126.
[11] H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian, "Exploiting enriched contextual information for mobile app classification," in Proc. 21stACMInt. Conf. Inform. Knowl. Manage., 2012, pp. 1617–1621.
[12] (2014). [Online]. Available: http://en.wikipedia.org/wiki/information_retrieval

Licensed Under Creative Commons Attribution CC BY