International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Fraud Detection of Mobile App Ranking

Rejin Idiculla Johnson[1], John Prakash[2]

[1]PG Scholar, Department of Computer Science & Engineering, Mangalore Institute of technology and engineering, Mangalore, Karnataka (India)

[2]Associate Professor, Department of Computer Science & Engineering, Mangalore Institute of technology and engineering, Mangalore, Karnataka (India)

Abstract: ***Fraud ranking among mobile app's, refers to the fraud activities that happens in the app market. These activities are aimed to bump up the apps in the popularity list. Some developers use shady means, like inflating sales of the app, or post fake app rating as ranking fraud. More popular apps will get more downloads and will result in developer getting more profit. So far there has been limited research in this field. This work is about developing a detection system for ranking fraud among mobile apps. The work is based on finding the leading sessions of mobile apps, which tells a time range in which the app is mostly used. The next step involves making use of three types of evidences. These are ranking, rating and review evidences. These will be then aggregated and compared with data collected from app datasets.***

Keywords: Ranking Fraud, App Rating and Review, Evidence Aggregation, Leading Session

1. Introduction

This is the age of mobile technology. People depend heavily on mobile devices. There exists 1.7 million apps as of now in Google's play store and Apples app store. Each apps will have a specific use and different interface. To promote the development of these mobile applications the different application stores also known as app markets has arranged a leaderboard chart. These apps are ranked in these charts based on their popularity.

An application which will have a higher ranking in the chart list will result in more number of downloads to it. This in turn will provide a source of revenue to the developer. The developer uses various marketing skills for the promotion. But since recent times, some developers are making use of fraud methods for promoting the apps. This will help in manipulation of chart ranking. The use of 'boosting' is such a method. These are software which will be tasked with keep on downloading an app thousands of times, hence resulting in improvement in its ranking. Another method is the use of 'human water armies'. These are paid writers in the internet who posts fake reviews for an application, thus making it appear a trusted product.

So far there has been limited development in the field of fraud detection of mobile app ranking. The work associated about this area comes from spam detection in online reviews and web spam detection. Due to the need of a proper detection mechanism, the proposed work is about a detection system for android mobile application. There are several factors which should be considered for this. One main factor is the time of popularity of an application. This is the time range in an application is popular and is widely used. This is also the time the application is most likely to be exposed to fraud activities. This time range is called a Leading Session. By identifying this leading session, the ranking pattern of the application can be determined. Further the app's rating and review details given by the users are also taken into account. These details can be compared against previous historical details of the application that has been collected. The historical details are obtained from datasets provided by Google's playstore.

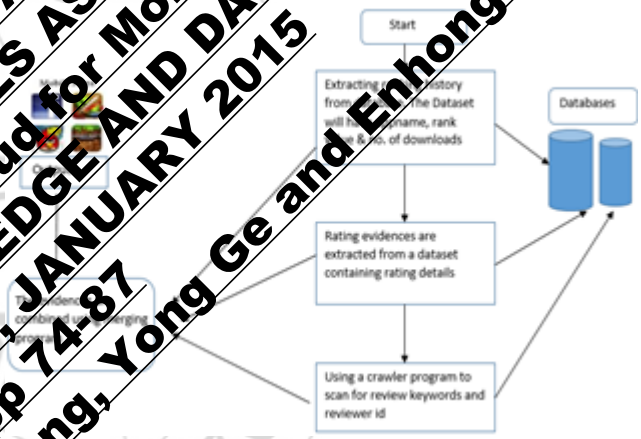


Figure 1: The Framework for the detection system

2. Related Work

The first is about web ranking spam detection. It refers to any deliberate actions which bring to selected webpages an unjustifiable favourable relevance or importance. For example, Ntoulaset have studied various aspects of content-based spam on the web and presented a number of Heuristic methods for detecting content based spam. Zhou have studied the problem of unsupervised webranking spam detection. Specifically, they proposed an efficient online link spam and term spam detection methods using spamicity. Recently, Spirin and Han have reported a survey on web spam detection, which comprehensively introduces the principles and algorithms in the literature. Actually, the work of web ranking spam detection is mainly based on the analysis of ranking principles of search engines, like PageRank and query term frequency. This is different from ranking fraud detection for mobile Apps.

The second category is concentrated on detecting online review spam. For example, Lim have identified several indicative behaviours of review spammers and model these behaviours to detect the spammers. Wu have studied the

problem of detecting hybrid shilling attacks on rating data. The proposed approach is based on the semi-supervised learning and can be used for reliable product recommendation. Xie have studied the problem ofsingleton review spam detection. Specifically, they solved this problem by detecting the co-anomaly patterns inmultiple review based time series. Although some of above 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.

The third category includes the studies on mobile App recommendation. For example, Yan and Chendeveloped a mobile App recommender system named Appjoy, which is based on user's App usage records to build a preference matrix instead of using explicit user ratings. Also, to solve the sparsity problem of App usage records, Shi and Ali studied several recommendation models and proposed a content based collaborative filter model called EigenApp. Some researchers studied the problem of exploiting enriched contextual information for personalized context aware recommendation which integrates both context dependency and independency assumptions.

3. Problem Statement and Proposed Solution

3.1 Problem Statement

An app with the higher ranking in the charts will have more downloads. Some app developers will use fraudulent means to deliberately boost their Apps. Some of the fraud activities will be usage of bot farms or human water armies. Due to the huge number of mobile Apps, it is difficult to manually label ranking fraud for each App. And due to the dynamic nature of chart rankings, it is not easy to identify and confirm the evidences linked to ranking fraud.

3.2 Proposed Solution

The proposed solution consists of the following phases.

3.3 Extracting Ranking Based Evidence

By analysing the Apps' historical ranking records, we observe that Apps' ranking behaviours in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, named rising phase, maintaining phase and recession phase. App's ranking first increases to a peak position in leader board, then keeping such peak position for a period and the ranking decreases till end of the event.

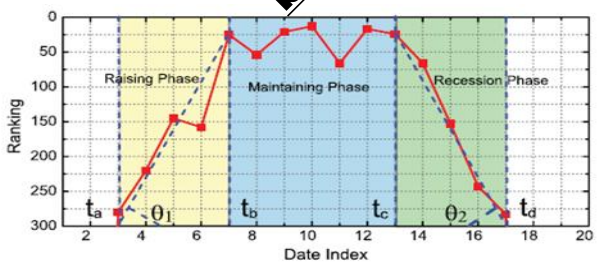


Figure 2: Ranking Pattern of an App

3.4 Extracting Rating Based Evidence

The rating evidence is obtained from a dataset containing apps historical details. The user ratings during a time period may have anomaly patterns compared with its historical rating. If an App has ranking fraud in a leading session, the ratings during the time period may have anomaly patterns compared with its historical ratings, which can be used for constructing rating based evidences

3.5 Extracting Review Based Evidence

Most of the App stores also permit users to write some textual comments as App reviews. Such reviews can indicates the individual perceptions and usage experiences of existing users for particular mobile Apps. Review manipulation is one of the most valuable perspectives 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 that contains more encouraging reviews may captivate more users to download. So imposters often posts fake reviews in the leading sessions of a particular App in order to increase the App downloads, and thus propel the App's ranking position in the leader board. The issue of detecting the local inconsistency of reviews in the leading sessions and capturing them as evidences for ranking fraud detection are still under explored. For this purpose, here we propose two fraud evidences for detecting ranking fraud based on Apps' review behaviours in leading session.

3.6 Evidence Aggregation

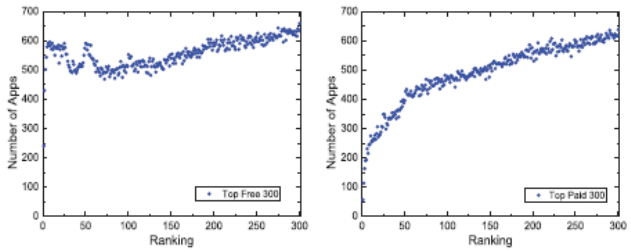
After extracting three types of fraud evidences, it is all combined using an unsupervised approach based on fraud similarity.

4. Experimental Result

Here the performance of the detection system is evaluated using real world App data.

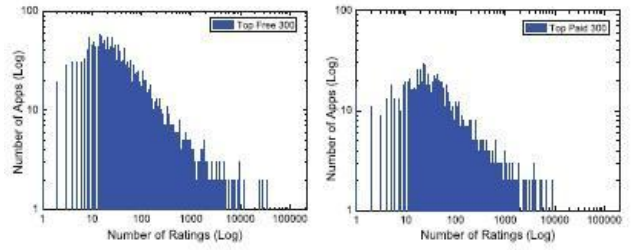
4.1 The Experimental Data

The experimental data sets were collected from the "Top Free 300" and "Top Paid 300" leader boards of Apple's App Store 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.3a and 3b indicate 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. 4a and 4b 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.



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

Figure 3(a) and **Figure 3(b)**



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

Figure 4(a) and **Figure 4(b)**

5. Conclusion

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. In this case, we identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. An optimization based aggregation method to integrate all the evidences for evaluating the risk of leading sessions from mobile Apps is proposed. All the evidences can be modelled by statistical hypothesis tests for the unique perspective of this approach, 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 Google play store. Experimental results showed the effectiveness of the proposed approach.

6. Acknowledgment

I take this opportunity to thank Prof. John Prakash, Prof. P V Bhat & Prof. DrNagaraj H R, for their valuable guidance and for providing all the necessary support to accomplish this research. I would like to extend my gratitude towards our beloved Principal G L Eshwar Prasad for his great support

References

[1] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proc. 19th ACM Int. Conf. Inform. Knowl. Manage 2010

[2] Z. Wu, J. Wu, J. Cao, and D. Tao, "HySAD: A semi-supervised hybrid shilling attack detector for trustworthy product recommendation," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012

[3] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012

[4] B. Zhou, J. Pei, and Z. Tang, "A spamicity approach to web spam detection," in Proc. SIAM Int. Conf. Data Mining, 2008

[5] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in Proc. 15th Int. Conf. World Wide Web, 2006

[6] B. Yan and G. Chen, "AppJoy: Personalized mobile application discovery," in Proc. 9th Int. Conf. Mobile Syst., Appl., Serv., 2011

[7] H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian, "Exploiting enriched contextual information for mobile app classification," in Proc. 21st ACM Int. Conf. Inform. Knowl. Manage, 2012

[8] H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian, "Mining personal context-aware preferences for mobile users," in Proc. IEEE 12th Int. Conf. Data Mining, 2012

[9] K. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach.Learn., 2007