[Kiruthika* 4(3): March, 2017] ISSN 2349-4506
Impact Factor: 2.785

Global Journal of Engineering Science and Research Management

A NOVEL DATA MINING APPROACH FOR RANKING FRAUD DETECTION USING AGGREGATION OF EVIDENCES

Kiruthika. N*
* PG Scholar in M.Sc Information Technology, PG Department of Information Technology, Hindusthan College of Arts and Science, Coimbatore

DOI: 10.5281/zenodo.376575

KEYWORDS: Applications, fraud detection, evidences, Historical Record.

ABSTRACT
Mobile application plays an important role for all the smart phone users to play or perform different tasks.Mobile application developers are available in large number; they can develop the different mobile applications. For making lager users for their applications some developers involve in illegal activities. Due to these illegal activities the mobile applications hires high rank in the application popularity list. Such fraudulent activities are used by more and more application developers. A ranking fraud detection system for mobile Apps is proposed in this paper. Accurately locate the ranking fraud by mining the leading sessions, of mobile Apps.R3-RFD algorithm is proposed in this paper. Furthermore, sentiword dictionary is used to identify the exact reviews scores. The fake feedbacks by a same person for pushing up that app on the leaderboard are restricted. Two different constraints are considered for accepting the feedback given to an application. The first constraint is that an app can be rated only once from a user login. And the second is implemented with the aid of MAC address that limits the number of user login logged per day from a MAC address as five.

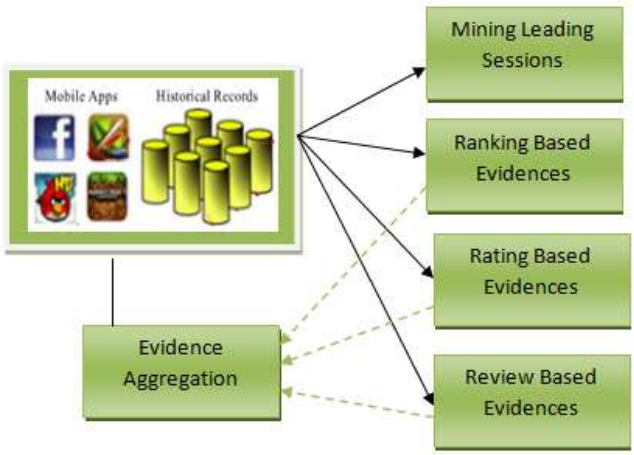
INTRODUCTION
The quantity of mobile Apps has developed in the course of recent years. The growth of apps was increased by 1.6 million. The top rated apps are displayed on the App leaderboards. To be sure, the App leaderboard is essential for improving the mobile Apps ratings and allowing the users to download the app. Ranking fraud in the mobile app market refers to illegal activities which have a purpose of boosting up the apps in the popularity list. To this end, in this paper, a novel approach for ranking fraud detection system for mobile apps is provided. Specifically, leading sessions, of mobile Appsis proposed to accurately locate the ranking fraud by mining the active periods.

PROBLEM STATEMENT
Many mobile app stores launched daily app leader boards which show the chart ranking of popular apps. The leader board is the important for promoting apps. Original application grade level decreases due to the duplication arrival in the mobile 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. User not understanding the Fake Apps then the user also gives the reviews in the fake application. Exact Review or Ratings or Ranking Percentage are not correctly Calculated.

Rank aggregation aims to combine multiple rankings ofitems (called also base rankers or ranking lists) generated byvarious sources (e.g. individual search engines) to produce abetter ranking. Several ranking fusion methodshave been proposed in the literature. These methods fallinto two categories: score-based aggregation and rank-basedaggregation. In the first category, items in the ranking listsare assigned scores and the ranking aggregation function usesthese scores in order to create the final list. In the secondcategory, items function relies only on the rank.

SYSTEM ARCHITECTURE
In recent years, mobile app has been rapidly growing while boosting more than 400,000 applications like Apple app store and Google Android market. This rapidimprovement of mobile App has made it complex to user for finding unique and trusted patterns of Application. Thus to solve this issue, marketing executives use rankingfor the App. In this paper, a useful R3-RFD algorithm is used to find the leading sessions and with the analysis of

those records, it is proved that apps usually have different ranking patterns in each sessions as compared to the normal apps. Therefore it is illustrated from those ranking records that some fraud is taking place in mobile app market and to restrict those frauds, evidences are developed to detect such fraud. As only ranking based evidences does not seems to be much sufficient to detect the fraud of mobile app, based on apps rating and review history some fraud evidences were discovered which showed anomaly patterns by those history. Specifically, an unsupervised evidence aggregation method is also proposed for evaluating the trustworthiness of leading sessions. And finally, the proposed system is estimated with real world app data gathered from the Google Play store for time consuming period. The results of these experiments showed an effectiveness of proposed approach in fig 1.



MODULES DESCRIPTION

Module 1: Leading events

Given a positioning limit K*∈ [1, K] a main occasion e of App a contains a period range also, relating rankings of a, Note that positioning edge K * is applied which is normally littler than K here on the grounds that K may be huge (e.g., more than 1,000), and the positioning records past K * (e.g., 300) are not exceptionally helpful for recognizing thepositioning controls. Moreover, it is finding that a few Apps have a few nearby driving even which are near one another and structure a main session.

Module 2: Identifying the important sessions for mobile apps

Basically, mining important sessions has two types of steps concerning with mobile fraud apps.Firstly, from the Apps ranking records, leading events is discovered and then secondly merging of adjacent important events is done. Theproposed algorithm is able to identify the certain leading events and the sessions by scanning records one by one.

Module 3: Identifying evidences for ranking fraud detection

Evidences Based on Ranking:

It concludes that important session comprises of various events. Hence by analysing the basic behaviour of leading events for finding fraud evidences and also for the app ranking records, it is been noted that a particular ranking pattern is always contained by app ranking behaviour in a leading event.

Evidences Based on Rating:

Previous ranking based techniques are useful for detection purpose but it is not efficient. As we know that rating is been done after downloading it by the user, and if the rating is high in leaderboard that is liked by most of the mobile app users. Spontaneously, the ratings in the important session gives rise to the unusual pattern which happens during rating fraud. These records can be used for creating evidences which is based on rating.

Evidences based on Review:
Mostly people tend to download Apps after reading the reviews. Therefore, due to some previous works on review spam detection, there still issue on locating the local anomaly of reviews in leading sessions. Based on apps review behaviours, fraud evidences are used to detect the Mobile app ranking fraud.

Advantages of proposed System.
1. Detect Fraud ranking in daily App leader boards.
2. Avoid ranking manipulation.

Mining Leading Sessions Algorithm

Algorithm 1 Mining Leading Sessions

Input 1: a's historical ranking records R_a;
Input 2: the ranking threshold K^*;
Input 2: the merging threshold ϕ;
Output: the set of a's leading sessions S_a;
Initialization: $S_a = \emptyset$;

1: $E_s = \emptyset$; $e = \emptyset$; $s = \emptyset$; $t^e_{start} = 0$;
2: for each $i \in [1, |R_a|]$ do
3: if $r^a_i \leq K^*$ and $t^e_{start} == 0$ then
4: $t^e_{start} = t_i$;
5: else if $r^a_i > K^*$ and $t^e_{start} \neq 0$ then
6: //found one event;
7: $t^e_{end} = t_{i-1}$; $e =< t^e_{start}, t^e_{end} >$;
8: if $E_s == \emptyset$ then
9: $E_s \cup = e$; $t^s_{start} = t^e_{start}$; $t^s_{end} = t^e_{end}$;
10: else if $(t^e_{start} - t^s_{end}) < \phi$ then
11: $E_s \cup = e$; $t^s_{end} = t^e_{end}$;
12: else then
13: //found one session;
14: $s =< t^s_{start}, t^s_{end}, E_s >$;
15: $S_a \cup = s$; $s = \emptyset$ is a new session;
16: $E_s = \{e\}$; $t^s_{start} = t^e_{start}$; $t^s_{end} = t^e_{end}$;
17: $t^e_{start} = 0$; $e = \emptyset$ is a new leading event;
18: return S_a

ISSN 2349-4506
Impact Factor: 2.785
Global Journal of Engineering Science and Research Management

SYSTEM DESIGN

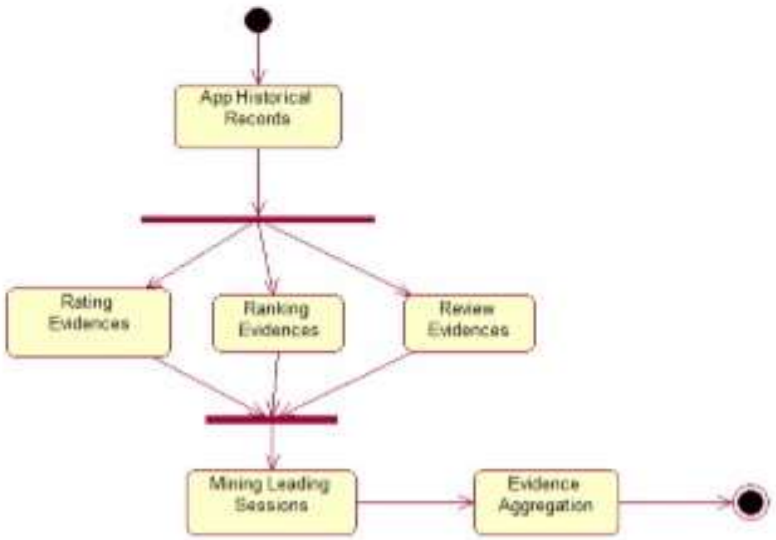


Fig 1: System flow design of fraud detection

EXPERIMENTAL RESULTS

Experimental Data

The experimental data sets were collected from the leaderboards of Google's Play Store.
The data sets contain the daily chart rankings of top free Apps and top paid Apps, respectively. The data set contains the ratings and reviews of the individual applications.

Two different constraints are considered for accepting the feedback given to an application. The first constraint is that an app can be rated only once from a user login. And the second is implemented with the aid of MAC address that limits the number of user login logged per day from a MAC address as five.

CONCLUSION

In this paper, the ranking fraud detection system for mobile applications is studied. For that purpose considered the most important sessions that are useful to identifying ranking fraud happened and gave a method for mining leading sessions. Then, identified all evidences like, ranking, rating and also review based evidences for detecting

fraud apps. After this, proposed a method which aggregates all these evidences for finding the fraud mobile applications.

The proposed system is estimated with real world app data gathered from the Google Play store for time consuming period.

REFERENCES

1. MAdFraud: Investigating Ad Fraud in Android Applications.
2. Mining Personal Context-Aware Preferences for Mobile Users.
3. A Flexible Generative Model for Preference Aggregation. [4] Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM '08, pages 219–230, 2008.
4. D. M. Blei, A. Y. Ng, and M. I. Jordan. Lantent dirichlet allocation. Journal of Machine Learning Research, pages 993–1022, 2003.
5. Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou. A taxi driving fraud detection system. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining, ICDM '11, pages 181–190, 2011.
6. N. Spirin and J. Han, "Survey on web spam detection: Principles and algorithms," SIGKDD Explor. Newslett., vol. 13, no. 2, pp. 50–64, May2012.
7. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proc.19thACMInt. Conf. Inform. Knowl. Manage., 2010, pp. 939–948.
8. Cay Horstmann, Wiley,"Big JAVA 3rd Edition, John Wiley & Sons, Inc(2008)
9. Herbert Schildt, Java 2: The Complete Reference, Fifth Edition, Tata McGraw Hill(2002).
10. Paul Dubois, MYSOL(4th Edition)", Addison Wesley(2008).
11. S. Chiasson, R. Biddle, and P. van Oorschot, A second look at the usability of click-based graphical passwords, in ACM Symposium on Usable Privacy and Security (SOUPS)(2007).
12. www.developer.com
13. www.devx.com
14. www.en.wikipedia.org
15. www.datamining.com