ISSN: 2455 - 9172

EIJO Journal of Engineering, Technology And Innovative Research (EIJO–JETIR)

Einstein International Journal Organization (EIJO)

Available Online at: www.eijo.in

Volume – 1, Issue – 2, May – June – 2016, Page No. : 01-04

Improved Hierarchial Ranking Fraud Detection of Mobile Apps Using Sentiword Dictonary

P.Ponnaruvi[1], P.Suganya[2], K. Raju[3], S. Sageengrana[4]

[1, 2, 3] Assistant professor, Department of Information Technology, E G S PILLAY Engineering College, Nagapattinam

[4]Assistant professor, Department of Computer Science, Vel Tech High Tech Dr.R.R Dr.S.R Engineering College, Avadi

E-Mail: [1]ponnaruvi@egspec.org, [2]Suganya.pasm@gmail.com, [3]profgkr@gmail.com, [4]granadhas@gmail.com

ABSTRACT

The primary aim of this project is to enhance the prevention of ranking frauds in mobile apps using the MAC address. In the existing system the leading event and leading session of an app is identified from the collected historical records. Then three different types of evidences are collected from the user feedbacks namely ranking based evidence, rating based evidence and review based evidence. These three evidences are aggregated by using evidence aggregation method. In the proposed system additionally, we are proposing two enhancements. Firstly, we are using Sentiword dictionary to identify the exact reviews scores. Secondly, the fake feedbacks by a same person for pushing up that app on the leader board 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.

Keywords: Ranking fraud, leader board, Evidence aggregation, MAC address, IP spoofing.

1. Introduction

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.

2. System Analysis

Many mobile app stores launched a daily app leader board which shows 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.

3. Implementation

A. Mobile Apps and Historical ranking records

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 & risqué; 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.

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 & rsquo; ranking behaviors, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps.

B. Ranking Fraud Detection

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.

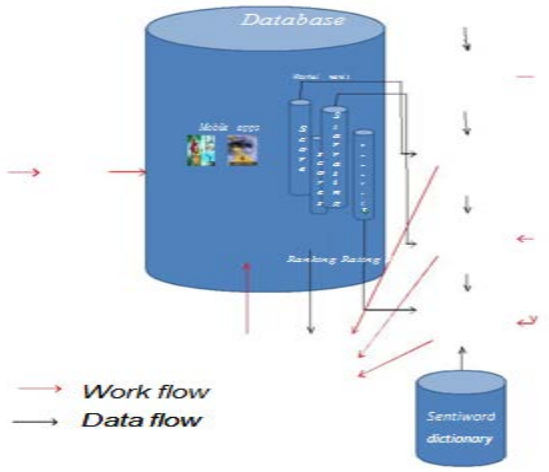

Fig 3.1: Ranking Fraud Detection of Mobile Apps

C. Rating and Review

We further propose two types of fraud evidences based on Apps & rsquo; rating and review history, rich reflect some anomaly patterns from Apps & rsquo; 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. Figure 3.1 shows the framework of our ranking fraud detection system for mobile Apps. It is worth noting that all the apps & nbsp; evidences are extracted by modeling Apps & rsquo; ranking, rating and review behaviors through statistical hypotheses tests. The proposed framework is scalable and can be extended with other domain generated evidences for ranking fraud detection.

D. Update App Details

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.

4. Evidence Aggregation

We develop an unsupervised evidence-aggregation method to integrate these three types of evidences for evaluating the credibility of leading sessions from mobile Apps. Figure 1 shows the framework of our ranking fraud detection system for mobile Apps.

A. ***MAC Address***

If the user gives ranking and rating many times for an app, then it will be identified by the admin using the MAC address. The user cannot give more than five ranking or rating for an app a day from one MAC. MAC address cannot be changed. Using MAC, IP spoofing attack can be blocked. The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

5. Algorithm

6. Mining Leading Session

There are two main steps for mining leading sessions. First, we need to discover leading events from the Apps historical ranking records. Second, we need to merge adjacent leading events for constructing leading sessions. Specifically, Algorithm 1 demonstrates the pseudo code of mining leading sessions for a given App a. Complaints of an original version of application provider can be undertaken by using Mining Leading Session algorithm. The duplicate version is identified by the admin by means of Historical Records. The admin will also see the date of publication of the apps. When the apps is detected as fraudulently published by the admin then the respective app will be blocked. The user can give the feedback at only once. Sentiword dictionary is used for finding the exact reviews. The admin can block the fake application. The Review or Rating or Ranking given by users is Correctly Calculated. Hence, a new user who wants to download an app for some purpose can get clear view about the available applications.

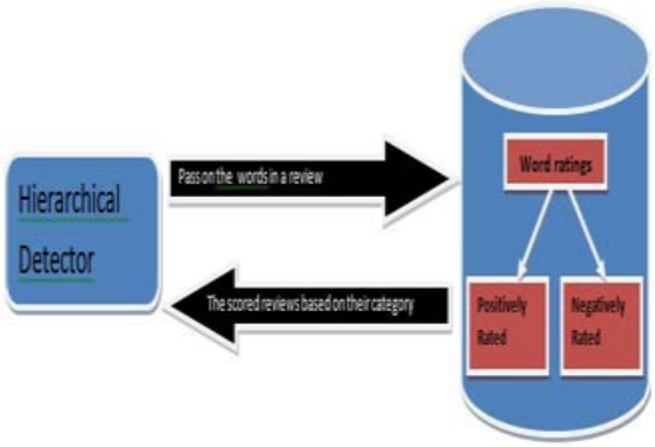

Fig 3.2: Usage of sentiword dictionary with hierarchical detector.

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

7. Conclusion

Hence the prevention of ranking frauds in mobile apps using the MAC address can be performed in an enhanced way with the help of integrating sentiword dictionary on with hierarchical method of detection. When compared to the existing system by using Sentiword dictionary exact reviews scores can be shortlisted apart from fake reviews for pushing up that app on the leader board and likewise fraudulent reviews for any other purpose apps for are restricted. This is been implemented by placing the following two restrictions such as restricting the review count allowed for a particular user to one and also with the help of MAC address allowing a particular MAC address to be used only once for accepting the feedback given to an application our expected level of throughput for detecting fraudulent apps is been increased.

8. References

[1]. Discovery of Ranking Fraud for Mobile Apps Hengshu Zhu, Hui Xiong, Senior Member, IEEE, Yong Ge, and Enhong Chen, Senior Member, IEEE.

[2]. Experiments with dbpedia, wordnet and sentiwordnet as resources for sentiment analysis in micro-blogging Hussam Hamdan, Frederic Béchet, Patrice Bellot Aggregation Process With Multiple Evidence Levels For Experimental Studies In Software Engineering Enrique Fernández.

[3]. L. Azzopardi, M. Girolami and K. V. Risjbergen, "Investigating the relationship between language model perplexity and ir precision-recall measures", Proc. 26th Int. Conf. Res. Develop. Inform. Retrieval, pp. 369-370, 2003 .

[4]. D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet allocation", J. Mach. Learn. Res., pp. 993-1022, 2003.

[5]. Y. Ge, H. Xiong, C. Liu and Z.-H. Zhou, "A taxi driving fraud detection system", Proc. IEEE 11th Int. Conf. Data Mining, pp. 181-190, 2011.

[6]. D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclear norm minimization", Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 60-68, 2011 .

[7]. T. L. Griffiths and M. Steyvers, "Finding scientific topics", Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228-5235, 2004.

[8]. G. Heinrich, Parameter estimation for text analysis, [online].

[9]. N. Spirin and J. Han, "Survey on web spam detection: Principles and algorithms", SIGKDD Explor. Newslett., vol. 13, no. 2, pp. 50-64, 2012 [CrossRef].

[10]. M. N. Volkovs and R. S. Zemel, "A flexible generative model for preference aggregation", Proc. 21st Int. Conf. World Wide Web, pp. 479-488, 2012 [CrossRef].

[11]. Z. Wu, J. Wu, J. Cao and D. Tao, "HySAD: A semi-supervised hybrid shilling attack detector for trustworthy product recommendation", Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 985-993, 2012 [CrossRef]

[12]. S. Xie, G. Wang, S. Lin and P. S. Yu, "Review spam detection via temporal pattern discovery",Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 823-831, 2012 [CrossRef].

[13]. B. Yan and G. Chen, "AppJoy: Personalized mobile application discovery", Proc. 9th Int. Conf. Mobile Syst., Appl., Serv., pp. 113-126, 2011 [CrossRef].

[14]. B. Zhou, J. Pei and Z. Tang, "A spamicity approach to web spam detection", Proc. SIAM Int. Conf. Data Mining, pp. 277-288, 2008 [CrossRef].

[15]. H. Zhu, H. Cao, E. Chen, H. Xiong and J. Tian, "Exploiting enriched contextual information for mobile app classification", Proc. 21st ACM Int. Conf. Inform. Knowl. Manage., pp. 1617-1621, 2012[CrossRef].

[16]. H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong and J. Tian, "Mining personal context-aware preferences for mobile users", Proc. IEEE 12th Int. Conf. Data Mining, pp. 1212-1217, 2012.