IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 12, December 2015

A Survey on Ranking Fraud Detection Using Opinion Mining for Mobile Apps

Tejaswini B. Gade[1], Prof. Nilesh G. Pardeshi[2]

PG Student, Computer Engineering Department, SRESCOE, Kopargaon, India[1]

Assistant Professor, Computer Engineering Department, SRESCOE, Kopargaon, India[2]

Abstract: Now days, mobile App is a very popular and well known concept due to the rapid advancement in the mobile technology and mobile devices. Due to the large number of mobile Apps, ranking fraud is the key challenge in front of the mobile App market. Ranking fraud refers to fraudulent or vulnerable activities which have a purpose of bumping up the Apps in the popularity list. In fact, it becomes more and more frequent for App developers to use tricky means, like increasing their Apps' sales or posting fake App ratings, to commit ranking fraud. While the importance and necessity of preventing ranking fraud has been widely recognized. After understanding the details of ranking fraud and the need of ranking fraud detection, the paper proposes a ranking fraud detection system for mobile Apps. The proposed system mines the active periods such as leading sessions of mobile apps to accurately locate the ranking fraud. These leading sessions can be useful for detecting the local anomaly instead of global anomaly of App rankings. Besides this, by modeling Apps ranking, rating and review behaviours using statistical hypotheses tests, we investigate three types of evidences, they are ranking based evidences, rating based evidences and review based evidences. Furthermore, we propose an aggregation method based on optimization to integrate all the evidences for fraud detection. Finally, the proposed system will be evaluated with real-world App data which is to be collected from the App Store for a long time period.

Keywords*:* Mobile Apps, ranking fraud detection, evidence aggregation, historical ranking records, rating and review.

I. INTRODUCTION

The number of mobile Apps has grown up on a very large scale over the past few years. Such as, there are more than 1.6 million Apps at Apple's App store and Google Play at the end of April 2013. To inspire the development of mobile Apps, many App stores launched daily App leader board, which shows the chart rankings of most popular Apps. This type of apps is the most important ways for promoting mobile Apps. A top rank on the leader board usually leads to a huge number of downloads and million dollars in revenue. As a result, App developers incline to explore various ways such as advertisement drive to promote their Apps to get higher position in such App leaderboards.

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, VentureBeat [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 leaderboard and more than 50,000-100,000 new users could be acquired within a couple of days. In actuality, such ranking fraud promotes great concerns to the mobile App industry. For example, Apple has notified of cracking down on App developers who commit ranking fraud [2] in the App store.

Leading events of mobile Apps forms different leading sessions. The mobile Apps not always ranked high in the leaderboards, but it usually happens in the leading sessions. So, detecting ranking fraud of mob Apps is actually the process to detect it within the leading session of the mobile Apps. Especially, this paper proposes a simple and effective algorithm to recognize the leading sessions of each mobile App based on its historical ranking records. This is one of the fraud evidence. Also, two types of fraud evidences are proposed based on Apps' rating and review history, which gives some anomaly patterns from Apps' historical rating and review records. In addition, we propose an unsupervised evidence-aggregation method to consolidate these three types of evidences for assessing the credibility of leading sessions from mobile Apps.

The rest of the paper is arranged as follows: Section II presents the literature survey over the related work. In section III, proposed system is presented. Finally, the section IV concludes the review paper.

II. LITERATURE SURVEY

In this section, we have studied previous research papers related to the detection of ranking fraud for mobile Apps. The research work of this study is divided into three categories. They are i) web ranking spam detection [3], [4], [5], ii) online review spam detection [6], [7], [8] and iii) mobile App recommendation [9], [10], [11], [12]. The first category is Web ranking spam detection. The web ranking spam refers to any intentional actions which bring to selected webpages an inexcusable auspicious relevant importance [5]. Following is the work done on web ranking spam detection.

DOI 10.17148/IJARCCE.2015.41278

A.Ntoulas *et al.* [3] presented a number of heuristic methods for detecting content based spam. He studied different aspects of content based spam on the web to find the heuristic methods.

N. Zhou *et al.* [5] studied the unsupervised web ranking spam detection. Using a spamicity, he proposed an efficient online link spam and spam detection methods.

Recently, B. Spirin *et al.* [4] has done a survey on Web spam detection. This survey thoroughly introduces the principles and algorithm in the literature. Certainly, the work of Web ranking spam is mainly based on the study of ranking principles of search engines, like page rank and query term frequency. This different from ranking fraud detection for mobile Apps.

Detection of ranking fraud for mobile Apps is still under a subject to research. To fill this crucial lack, we propose to develop a ranking fraud detection system for mobile Apps. We also determine several important challenges. First challenge, in the whole life cycle of an App, the ranking fraud does not always happen, so we need to detect the time when fraud happens. This challenge can be considered as detecting the local anomaly in place of global anomaly of mobile Apps. Second challenge, it is important to have a scalable way to positively detect ranking fraud without using any basis information, as there are huge number of mobile Apps, it is very difficult to manually label ranking fraud for each App. Finally, due to the dynamic nature of chart rankings, it is difficult to find and verify the evidences associated with ranking fraud, which motivates us to discover some implicit fraud patterns of mobile Apps as evidences.

III.PROPOSED SYSTEM

With the increase in the number of web Apps, to detect the fraudulent Apps, we have propose a simple and effective algorithm which identifies the leading sessions of each App based on its historical ranking of records. By analysing the ranking behaviours of Apps, we discover that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, we identify some fraud evidences from Apps' historical ranking records and develop three functions to obtain such ranking based fraud evidences.

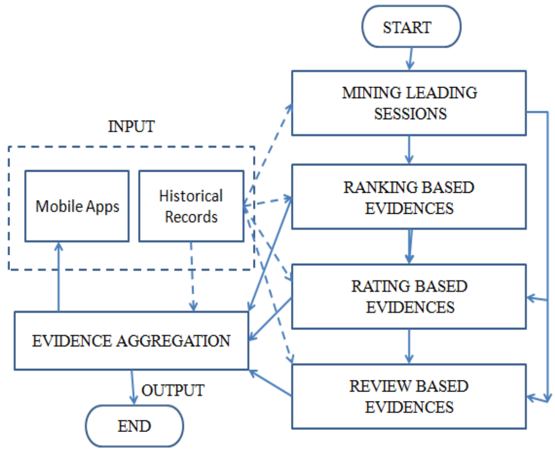

Fig. 1. Ranking fraud detection system framework

Further, we propose two types of fraud evidences based on Apps' rating and review history. It reflects some anomaly patterns from Apps' historical rating and review records. Fig. 1 shows the framework of our ranking fraud detection system for mobile Apps.

The leading sessions of mobile App signify the period of popularity, and so these leading sessions will comprise of ranking manipulation only. Hence, the issue of identifying ranking fraud is to identify vulnerable leading sessions. Along with this, the main task is to extract the leading sessions of a mobile App from its historical ranking records.

There are two main phases for detecting the ranking fraud:
i) Identifying the leading sessions for mobile apps
ii) Identifying evidences for ranking fraud detection

Let us see them in brief

A. *Identifying the leading sessions for mobile apps*

Primarily, mining leading sessions has two types of steps concerning with mobile fraud apps. First, from the Apps historical ranking records, discovery of leading events is done and then second merging of adjacent leading events is done which appeared for constructing leading sessions. Certainly, some specific algorithm is demonstrated from the pseudo code of mining sessions of given mobile App and that algorithm is able to identify the certain leading events and sessions by scanning historical records one by one.

B. *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, it is been observed that a specific ranking pattern is always satisfied by app ranking behaviour in a leading event

2) Rating based evidences:

Previous ranking based evidences are useful for detection purpose but it is not sufficient. Resolving the "restrict time depletion" problem, fraud evidences recognition is planned due to app historical rating records. 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. Spontaneously, the ratings during the leading session gives rise to the anomaly pattern which happens during rating fraud. These historical records can be used for developing rating based evidences.

3) Review based evidences:

We are familiar with the review which contains some textual comments as reviews by app user and before downloading or using the app user mostly prefer to refer the reviews given by most of the users. Therefore, although due to some previous works on review spam detection [13] there still issue on locating the local anomaly of reviews in leading sessions. So based on apps review behaviors, fraud evidences are used to detect the ranking fraud in Mobile App.

These three evidences will be integrated by an unsupervised evidence-aggregation method for evaluating the credibility of leading sessions from mobile Apps. The statistical hypotheses tests models Apps' ranking, rating and review behaviors to extract all the evidences. The ranking fraud detection framework is scalable and can be extended with other domain generated evidences for ranking fraud detection. Finally, we will evaluate the proposed system with real-world App data collected from the Apple's App store for a long time span, i.e., more than two years.

IV. CONCLUSION

This paper reviews various existing methods used for web spam detection, which is related to the ranking fraud for mobile Apps. Also, we have seen references for online review spam detection and mobile App recommendation.

By mining the leading sessions of mobile Apps, we aim to locate the ranking fraud. The leading sessions works for detecting the local anomaly of App rankings. The system aims to detect the ranking frauds based on three types of evidences, such as ranking based evidences, rating based evidences and review based evidences. Further, an optimization based aggregation method combines all the three evidences to detect the fraud.

ACKNOWLEDGEMENT

We are glad to express our sentiments of gratitude to all who rendered their valuable guidance to us. We would like to express our appreciation and thanks to the Principal of our college. We are also thankful to the Head of Department. We thank to the anonymous reviewers for their valuable comments.

REFERENCES

[1] (2012). [Online]. Available: http://venturebeat.com/2012/07/03/apples-crackdown-on-app-ranking-manipulation/

[2] (2012). [Online]. Available: https://developer.apple.com/news/index.php?id=02062012a

[3] 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, pp. 83–92.

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

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

[6] 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.

[7] 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, pp. 985–993.

[8] 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, pp. 823–831.

[9] K. Shi and K. Ali, "Getjar mobile application recommendations with very sparse datasets," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 204–212.

[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. 21st ACMInt. Conf. Inform. Knowl. Manage., 2012, pp. 1617–1621.

[12] 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, pp. 1212–1217.

[13] G. Shafer, A Mathematical Theory of Evidence. Princeton, NJ, USA: Princeton Univ. Press, 1976.

DOI 10.17148/IJARCCE.2015.41278