International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Survey on Fraud Ranking in Mobile Apps

Monali Zende[1], Aruna Gupta[2]

[1]Department of Computer Engineering, Savitribai Phule Pune University, Pune, India

[2]Professor, Department of Computer Engineering, Savitribai Phule Pune University, Pune, India

Abstract: ***Ranking fraud in the mobile App business suggest to false or tricky exercises which have a motivation behind, knocking up the Apps in the fame list. Now a days, many shady means are used more frequently by app developers, such expanding their Apps' business or posting imposter App evaluations, to confer positioning misrepresentation. There is a limited understanding and research area for preventing ranking fraud. This paper gives a whole perspective of positioning misrepresentation and describes a Ranking fraud identification framework for mobile Apps. This work is grouped into three category. First is web ranking spam detection, second is online review spam detection and last one is mobile app recommendation. The Web ranking spam refers to any deliberate actions which bring to selected Web pages an unjustifiable favorable relevance or importance. Review spam is designed to give unfair view of some products so as to influence the consumers' perception of the products by directly or indirectly influating or damaging the product's reputation.***

Keywords: Mobile Apps, Ranking Fraud Detection, Evidence Aggregation, Historical Ranking Records, Rating and Review.

1. Introduction

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 leaderboards, which demonstrate the chart rankings of most popular Apps. Indeed, the App leaderboard is one of the most important ways for promoting mobile Apps. A higher rank on the leaderboard usually leads to a huge number of downloads and million dollars in revenue. 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 leaderboards. 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. This is usually implemented by using so called "*bot farms*" or "*human water armies*" to inflate the App downloads, ratings and reviews in a very short time[10].

There are some related works, for example, web positioning spam recognition, online survey spam identification and portable App suggestion, but the issue of distinguishing positioning misrepresentation for mobile Apps is till under-investigated. The problem of detecting ranking fraud for mobile Apps is still underexplored. To overcome these essentials, in this paper, we build a system for positioning misrepresentation discovery framework for portable apps that is the model for detecting ranking fraud in mobile apps. For this, we have to identify several important challenges. First, fraud is happen any time during the whole life cycle of app, so the identification of the exact time of fraud is needed. Second, due to the huge number of mobile Apps, it is difficult to manually label ranking fraud for each App, so it is important to automatically detect fraud without using any basic information.

Mobile Apps are not always ranked high in the leaderboard, but only in some *leading events* ranking that is fraud usually happens in leading sessions. Therefore, main target is to detect ranking fraud of mobile Apps within leading sessions. 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 characterize from Apps' historical ranking records. Then three functions are developed to extract such ranking based fraud evidences. Therefore, further two types of fraud evidences are proposed based on Apps' rating and review history, which reflect some anomaly patterns from Apps' historical rating and review records. In addition, to integrate these three types of evidences, an unsupervised evidence-aggregation method is developed which is used for evaluating the credibility of leading sessions from mobile Apps.

2. Related Work

The related works of this study is grouped into three categories. The first category is about Web ranking spam detection. Specifically, the Web ranking spam refers to any deliberate actions which bring to selected Web pages an unjustifiable favorable relevance or importance. In this, the problem of unsupervised web spam detection is studied. They introduce the concept of spamicity to measure how likely a page is spam. Spamicity is more flexible and user-controllable measure than the traditional supervised classification methods. They propose efficient online link spam and term spam detection methods using spamicity. This methods do not need training and also cost effective. A real data set is used to evaluate the effectiveness and the efficiency [1].

For example, Ntoulas *et al.* [2] have studied various aspects of content-based spam on the Web and presented a number of heuristic methods for detecting content based spam.

In this paper, they continue investigations of "web spam": the injection of artificially-created pages into the web in order to influence the results from search engines, to drive traffic to certain pages for fun or profit. This paper considers some previously-undescribed techniques for automatically detecting spam pages, examines the effectiveness of these techniques in isolation and when aggregated using classification algorithms.

Zhou *et al* [1] have studied the problem of unsupervised Web ranking spam detection. Specifically, they proposed an efficient online link spam and term spam detection methods using spamicity.

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 behaviour, clicks, HTTP sessions. In turn, there is a subcategorization 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.

The second category is focused on detecting online review spam. For example, Lim *et al.* [4] have identified several representative behaviors of review spammers and model these behaviors to detect the 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.

Wu *et al.* [5] 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 trustworthy product recommendation. This paper presents a Hybrid Shilling Attack Detector, or HySAD for short, to tackle these problems. In particular, HySAD introduces MC-Relief to select effective detection metrics, and Semi-supervised Naive Bayes (SNBλ) to precisely separate Random-Filler model attackers and Average-Filler model attackers from normal users.

Xie *et al.* [6] have studied the problem of singleton review spam detection. Specifically, they solved this problem by detecting the co-anomaly patterns in multiple 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 (i.e., leading session).

Finally, the third category includes the studies on mobile App recommendation. For example, Yan *et al.* [7] developed 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 *et al.* [8] studied several recommendation models and proposed a content based collaborative filtering model, named Eigenapp, for recommending Apps in their Web site Getjar. In addition, some researchers studied the problem of exploiting enriched contextual information for mobile App recommendation.

For example, Zhu *et al.* [9] proposed a uniform framework for personalized context-aware recommendation, which can integrate both context independency and dependency assumptions. However, to the best of our knowledge, none of previous works has studied the problem of ranking fraud detection for mobile App**S.** Summary about the related work is given in below Table1 as Survey Table:

Table 1: Survey Table

Sr.no	Paper	Technique	Advantage	Disadvantage	Result
1	A Spamicity Approach to Web Spam Detection [1]	Link spamicity	do not need training	Can not find topical spam patterns	Effective and efficient to detect spam pages.
2	Detecting Spam Web Pages through Content Analysis [2]	Content based spam detection algorithm	classifier can correctly identify 86.2% of all spam pages		detecting contentbased spam
3	Detecting Product Review Spammers using Rating Behaviors [4]	Rating behavioral approach to detect review spammers	detect users generating spam reviews or review spammers	cannot incorporate review spammer detection into review detection and vice versa.	show that the detected spammers have more significant impact on ratings compared with the unhelpful reviewers.
4.	HySAD: A Semi-Supervised Hybrid Shilling Attack Detector for Trustworthy Product Recommendation [5]	Hybrid Shilling Attack Detector	precisely separate Random-Filler model attackers and Average-Filler model attackers from normal users and effective against hybrid attacks		effectively improve the accuracy of a collaborative-filtering based recommender system, and provide interesting opportunities for in-depth analysis of attacker behaviors.
5.	Review Spam Detection via Temporal Pattern Discovery [6]	hierarchical algorithm	effective in detecting singleton review attacks.		robustly detect the time windows where such attacks are likely to have happened.

3. Architectural View

With the increase in the number of web Apps, to detect the fraud Apps, this paper propose a simple and effective system. Fig.1 shows the Framework of Fraud ranking discovery in mobile app

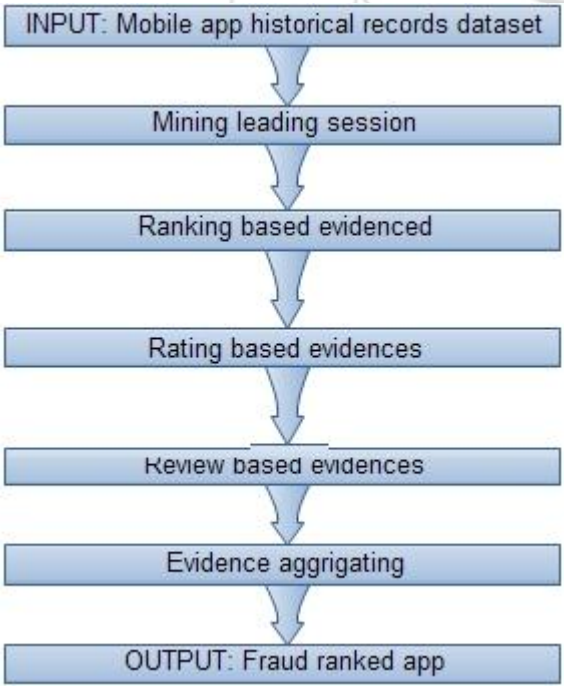

Figure1: Framework of Fraud ranking discovery in mobile app.

1. Module 1: Leading events

Given a positioning limit K _ 2 [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 the positioning 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.

2. Module 2: Leading Sessions

Instinctively, mainly 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 deceptive leading sessions. Along with the main task is to extract the leading sessions of a mobile App from its historical ranking records.

3. Module 3: Identifying the leading sessions for mobile apps

Basically, mining leading sessions has two types of steps concerning with mobile fraud apps.Firstly, from the Apps historical ranking records, discovery of leading events is done and then secondly 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.

4. Module 4: Identifying evidences for ranking fraud detection

4.1 Ranking Based Evidence

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.

4.2 Rating Based Evidence

Previous ranking based evidences are useful for detection purpose but it is not sufficient. Resolving the problem of "restrict time reduction", identification of fraud evidences 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.

4.3 Review Based Evidence

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, 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.

4. Conclusion

This paper, gives the ranking fraud detection model for mobile apps. Now a days many of mobile app developers uses various frauds techniques to increase their rank. To avoid this, there are various fraud detection techniques which are studied in this paper. We detect the ranking fraud using actual fraud reviews. This paper propose the time efficient system to detect the fraud Apps.

References

[1] 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.
[2] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly. Detecting spam web pages through content analysis. In *Proceedings of the 15th international conference on World Wide Web*, WWW '06, pages 83–92, 2006.
[3] N. Spirin and J. Han. Survey on web spam detection: principles and algorithms. *SIGKDD Explor. Newsl.*, 13(2):50–64, May 2012.
[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.
[5] 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
[6] 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.
[7] B. Yan and G. Chen. Appjoy: personalized mobile application discovery. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, MobiSys '11, pages 113–126, 2011.
[8] K. Shi and K. Ali. Getjar mobile application recommendations with very sparse datasets. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pages 204–212, 2012.
[9] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in VLDB, 1994.
[10] H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian. Mining personal context-aware preferences for mobile users. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 1212–1217, 2012.
[11] Hengshu Zhu, Hui XiongDiscovery of Ranking Fraud for Mobile Apps. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,2013.