International Journal of Research
Available at https://edupediapublications.org/journal

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 12
August 2016

A Novel Approach to Discovery of Ranking Fraud for Mobile Apps

Poornima
PG Scholar
Department of CSE,
DIET, ANAKAPALLE, Visakhapatnam

Dinesh Chandrasehkaran
Sr. Assistant Professor,
Department of CSE,
DIET, ANAKAPALLE, Visakhapatnam

Abstract - *Now a day's 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 therapy' sales or posting phony App ratings, to commit ranking fraud. The importance of preventing ranking fraud has been widely recognized. 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. Finally, we evaluate the proposed system with real-world App data collected from the iOS App Store for a long time period. In experiments, we validate the effectiveness of the proposed system, and shows the scalability of the detection algorithm as well assume regularity of ranking fraud activities.*

Key Words: ***Apps, ranking fraud detection, evidence aggregation, historical ranking records, rating and review*****.**

1. INTRODUCTION

The number of mobile apps has grown at a breathtaking rate over the past few years. For example, as of the end of april 2013, there are more than 1.6 million apps at apple's app store and google play. To stimulate the development of mobile apps, many app stores launched dailyapp 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. 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 "botfarms" or "human water armies" to inflate the app downloads ratings and reviews in a very short time. For example, an article from venture beat reported that, when an app was promoted with the help of ranking manipulation, it. In the literature, while there is some related work, such as web ranking spam detection online review spam detection and mobile App recommendation the problem of detecting ranking fraud for mobile Apps is still under-explored. To fill this crucial void, in this paper, we propose to develop a ranking fraud detection system for mobile Apps. Along this line, we identify several important challenges. First, ranking fraud does nodal ways happen in the whole life cycle of an App, so we need to detect the time when fraud happens. Such challenge can be regarded as detecting the local anomaly instead of global anomaly of mobile Apps. 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 have a scalable way to automatically detect ranking fraud without using any bench mark in formation. Finally, due to the dynamic nature of chart rankings, it is not easy to identify and confirm the evidences linked to ranking fraud, which motivates us to discover some implicit fraud patterns of mobile Apps as evidences. Indeed, our careful observation reveals that mobile Apps are not always ranked high in the leader board, but only in some leading events, which form different leading sessions. Note that we will introduce both leading events and leading sessions in detail later. 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.

2. RELATED WORK

This paper aims to detect users generating spam reviews or review spammers. We identify several characteristic behaviors of review spammers and model these behaviors so as to detect the spammers. In particular, we 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 reviewer in their ratings of products. We propose scoring methods to measure the degree of spam for each reviewer and apply them on an Amazon review dataset. We then select a subset of highly suspicious reviewers for further scrutiny by our user evaluators with the help of a web based spammer evaluation software specially developed for user evaluation experiments. Our results show that our proposed ranking and supervised methods are effective in discovering spammer sand outperform other baseline method based on helpfulness votes alone. We finally show that the detected spammers have more significant impact on ratings compared with the unhelpful reviewers.

From this paper we have referred:

- Concept of extracting of rating and ranking.
- Concept of extracting of review.

Advances in GPS tracking technology have enabled us to install GPS tracking devices in city taxis to collect a large amount of GPS traces under operational time constraints. These GPS traces provide unparalleled opportunities for us to uncover taxi driving fraud activities. In this paper, we develop a taxi driving fraud detection system, which is able to systematically investigate taxi driving fraud. In this system, we first provide functions to find two aspects of evidences: travel route evidence and driving distance evidence. Furthermore, a third function is designed to combine the two aspects of evidences based on dempster Shafer theory. To implement the system, we first identify interesting sites from a large amount of taxi GPS logs. Then, we propose a parameter free method to mine the travel route evidences. Also, we introduce route mark to represent a typical driving path from an interesting site to another one. Based on route mark, we exploit a generative statistical model to characterize the distribution of driving distance and identify the driving distance evidences. Finally, we evaluate the taxi driving fraud detection system with large scale real-world taxi GPS logs. In the experiments, we uncover some regularity of driving fraud activities and investigate the motivation of drivers to commit a driving fraud by analyzing the produced taxi fraud data. Concept of fraud detection Evaluative texts on the Web have become a valuable source of opinions on products, services, events, individuals, etc. Recently, many researchers have studied such opinion sources as product reviews, forum posts, and blogs. However, existing research has been focused on classification and summarization of opinions using natural language processing and data mining techniques. An important issue that has been neglected so far is opinion spam or trustworthiness of online opinions. In this paper, we study this issue in the context of product reviews, which are opinion rich and are widely used by consumers and product manufacturers. In the past two years, several startup companies also appeared which aggregate opinions from product reviews. It is thus high time to study spam in reviews. To the best of our knowledge, there is still no published study on this topic, although Web spam and email spam have been investigated extensively. We will see that opinion spam is quite different from Web spam and email spam, and thus requires different detection techniques. Based on the analysis of 5.8 million reviews and 2.14 million reviewers from amazon.com, we show that opinion spam in reviews is widespread. This paper analyzes such spam activities and presents some novel techniques to detect them. Many applications in information retrieval, natural language processing, data mining, and related fields require a ranking of instances with respect to specified criteria as opposed to a classification. Furthermore, for many such problems, multiple established ranking models have been well studied and it is desirable to combine their results into a joint ranking, formalism denoted as rank Aggregation. This work presents a novel unsupervised learning algorithm for rank aggregation (ULARA) which returns a linear combination of the individual ranking functions based on the principle of rewarding ordering agreement between the rankers. In addition to presenting ULARA, we demonstrate its effectiveness on a data fusion task across ad hoc retrieval systems.

3. EXISTING SYSTEM

In the literature, while there are some related work, such as web ranking spam detection, online review spam detection and mobile App recommendation, the problem of detecting ranking fraud for mobile Apps is still under-explored. Generally speaking, the related works of this study can be grouped into three categories. The first category is about web ranking spam detection. The second category is focused on detecting online review spam. Finally, the third category includes the studies on mobile App recommendation. This is different from ranking fraud detection for mobile Apps. The second category is focused on detecting online review spam have identified several representative behaviors of review spammers and model these behaviors to detect the spammers. We 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.

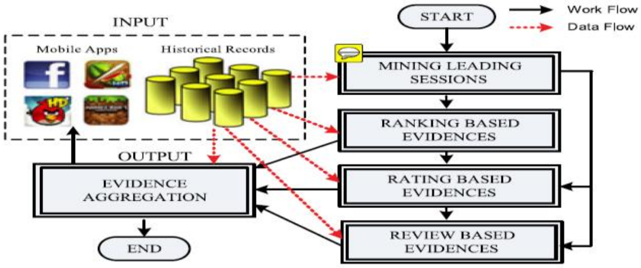
Xie have studied the problem of singleton review spam detection. Specifically, they solved this problem by detecting the co-anomaly patterns in multiple review based time series. Although some of above approaches can be used for anomaly detection from historical rating and review records, they are not 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. 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 content based collaborative filtering model, named Eigenapp, for recommending Apps in their website Getjar. In addition, some researchers studied the problem of exploiting enriched contextual information for mobile App recommendation. We 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 Apps.

3. 1 DISADVANTAGES OF EXISTING SYSTEM

- Although some of the existing 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).
- Cannot able to detect ranking fraud happened in Apps' historical leading sessions
- There is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud.

4. PROPOSED SYSTEM

In proposed system we overcome the drawbacks of Mining leading session algorithm which is based on ranking, review & rating. First, the download information is an important signature for detecting ranking fraud, since ranking manipulation is to use so-called "bot farms" or "human water armies" to inflate the App download and ratings in a very short time. However, the instant download information of each mob. App is often not available for analysis. In fact, Apple and Google do not provide accurate download information on any App. Furthermore, the App developers themselves are also reluctant to release their download information for various reasons. Therefore, in this paper, the focus is on extracting evidences from Apps' historical ranking, rating and review records for ranking fraud detection. However, our approach is scalable for integrating other evidences if available, such evidences based on the download information and App developers' reputation. Second, the proposed approach can detect ranking fraud happened in A,' historical leading sessions. Ranking fraud detection in mobile apps is actually to detect ranking fraud within leading session of mobile apps. Specifically we identified first leading sessions based on Apps historical ranking records. Then with the analysis of Apps'ranking behaviors we characterized some fraud evidences from historical records.



The ranking based evidences can be affected some Apps'developer reputation and some legitimate marketing campaigns, such as ―limited-time discount‖. This method is not enough to detect fraudulent Apps' so we propose two new methods of fraud evidences based on Apps' historical rating and review records. Additionally, we developed an unsupervised evidence-aggregation method to integrate these types of evidences. 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' ranking behaviors, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, we characterize some fraud evidences from Apps' historical ranking records, and develop three functions to extract such ranking based fraud evidences.

We further propose two types of fraud evidences based on Apps' rating and review history, which reflect some anomaly patterns from Apps' historical rating and review records.

In Ranking Based Evidences, by analyzing the Apps' historical ranking records, we observe that Apps' ranking behaviors in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase.

In Rating Based Evidences, specifically, after an App has been published, it can be rated by any user who downloaded it. Indeed, user rating is one of the most important features of App advertisement. An App which has higher rating may attract more users to download and can also be ranked higher in the leaderboard. Thus, rating manipulation is also an important perspective of ranking fraud.

In Review Based Evidences, besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most important perspectives of App ranking fraud.

4. 1 ADVANTAGES

- The proposed framework is scalable and can be extended with other domain generated evidences for ranking fraud detection.
- Experimental results show the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity of ranking fraud activities.
- To the best of our knowledge, there is no existing benchmark to decide which leading sessions or Apps really contain ranking fraud. Thus, we develop four intuitive baselines and invite five human evaluators to validate the effectiveness of our approach Evidence Aggregation based Ranking Fraud Detection (EA-RFD).

5. EXPERIMENT RESULT

To start with the mining driving sessions is utilized to find driving occasions from the application's chronicled positioning records and after that it blends nearby driving occasions for building driving sessions. At that point the positioning based proofs dissect the fundamental attributes of driving occasions for separating misrepresentation confirmations.

Methodology with relevancy every objective is concisely given below:

To provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile Apps. First propose to accurately find the ranking fraud by mining the active periods, specifically leading sessions, of mobile Apps.

Identifying leading sessions for mobile apps:

During this section, we have a tendency to 1st introduce some preliminaries, so show a way to mine leading sessions for mobile Apps from their historical ranking records.

Mining Leading Sessions

There ar 2 main steps for mining leading sessions. First, we'd like to find leading events from the App's historical ranking records. Second, we'd like to merge adjacent leading events for constructing leading sessions

To improve the fraud detection efficiency

Mining Leading Sessions

In the first module, we develop our system environment with the details of App like an app store. Intuitively, the leading sessions of a mobile App represent its periods of popularity, so the ranking manipulation will only take place in these leading sessions. Therefore, the problem of detecting ranking fraud is to detect fraudulent leading sessions. Along this line, the first task is how to mine the leading sessions of a mobile App from its historical ranking records. There are two main steps for mining leading sessions. First, we need to discover leading events from the App's historical ranking records. Second, we need to merge adjacent leading events for constructing leading sessions.

Ranking Based Evidences

In this module, we develop Ranking based Evidences system. By analyzing the Apps' historical ranking records, web serve that Apps' ranking behaviors in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase. Specifically, in each leading event, an App's ranking first increases to a peak position in the leaderboard (i.e., rising phase), then keeps such peak position for a period (i.e., maintaining phase), and finally decreases till the end of the event (i.e., recession phase).

Rating Based Evidences

In the third module, we enhance the system with Rating based evidences module. The ranking based evidences are useful for ranking fraud detection.

However, sometimes, it is not sufficient to only use ranking based evidences. For example, some Apps created by the famous developers, such as Gameloft, may have some leading events with large values of u1 due to the developers' credibility and the "word-of-mouth" advertising effect. Moreover, some of the legal marketing services, such as "limited-time discount", may also result in significant ranking based evidences. To solve this issue, we also study how to extract fraud evidences from Apps' historical rating records.

Review Based Evidences

In this module we add the Review based Evidences module in our system. Besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most important perspectives of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users often first read its historical reviews to ease their decision making, and a mobile App contains more positive reviews may attract more users to download. Therefore, imposters often post fake reviews in the leading sessions of a specific App in order to inflate the App download, and thus propel the App's ranking position in the leader board.

Evidence Aggregation

In this module we develop the Evidence Aggregation module to our system. After extracting three types of fraud evidences, the next challenge is how to combine them for ranking fraud detection. Indeed, there are many ranking and evidence aggregation methods in the literature, such as permutation based models score based models and Dempster-Shafer rules . However, some of these methods focus on learning a global ranking for all candidates. This is not proper for detecting ranking fraud for new Apps. Other methods are based on supervised learning techniques, which depend on the labeled training data and are hard to be exploited. Instead, we propose an unsupervised approach based on fraud similarity to combine these evidences.

6. CONCLUSION

In this paper, 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. Moreover, we proposed an mining Leading session algorithm for obtain mining leading session and aggregation method. In the future, we plan to study more effective fraud evidences and analyze the latent relationship among rating, review and rankings. Moreover, we will extend our ranking fraud detection approach with other mobile App related services, such as mobile Apps recommendation, for enhancing user experience. Provide a holistic read of ranking fraud and propose a ranking fraud detection system for mobile Apps. 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. Moreover, we proposed an optimization based aggregation method to integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps. First propose to accurately find the ranking fraud by mining the active periods, particularly leading sessions of mobile Apps. For achieving this goal following process are used: Distinctive leading sessions for mobile apps during this section, we tend to 1st introduce some preliminaries, so show a way to mine leading sessions for mobile Apps from their historical ranking records. First, we'd like to get leading events from the App's historical ranking records. Second, we'd like to merge adjacent leading events for constructing leading sessions. We extended our ranking fraud detection approach with other mobile app related services, such as mobile app recommendation for enhancing user experience.

REFERENCES

[1] (2014). http://en.wikipedia.org/wiki/cohen's kappa

[2] (2014). http://en.wikipedia.org/wiki/inform retrieval

[3](2012).https://developer.apple.com/index.php?id=02062012a

[4] (2012). http://venturebeat.com/2012/07/03/apples-crackdown-on-app-ranking-manipulation/

[5] (2012). http://www.ibtimes.com/applethreatens-crackdown-biggest-app-store-ranking-fra ud-406764

[6] (2012). http://www.lextek.com/manuals/index.html

[7] (2012).
http://www.ling.gu.se/lager/mogul/porter-stemmer.

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

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

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

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

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

[13] G. Heinrich, Parameter estimation for text analysis, " Univ. Leipzig, Leipzig, Germany, Tech. Rep., http://faculty.cs.byu.edu/~ringger/CS601R/papers/Heinrich-GibbsLDA.pdf, 2008.

[14] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int.Conf. Web Search Data Mining, 2008, pp. 219–230.

[15] J. Kivinen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," in Proc. 27th Annu. ACM Symp. Theory Comput., 1995, pp. 209–218.

[16] A. Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach.Learn., 2007, pp. 616–623.

[17] A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with distance-based models," in Proc. 25th Int. Conf. Mach. Learn., 2008, pp. 472–479.

[18] A. Klementiev, D. Roth, K. Small, and I. Titov, "Unsupervised rank aggregation with domain-specific expertise," in Proc. 21st Int. Joint Conf. Artif. Intell., 2009, pp. 1101–1106.

[19] 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.

[20] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li, "Supervised rank aggregation," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 481–490.

[21] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 632–640.

[22] 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.
[23] G. Shafer, A Mathematical Theory of Evidence. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[24] 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.

[25] 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.

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

BIOGRAPHIES

POORNIMA,
STUDENT OF M.TECH.,
DADI INSTITUTE OF ENGG.&TECH.,
ANAKAPALLI,
ANAKAPALLI. ,A.P,INDIA.

C.DINESH , M.TECH(C.S.E),
SR.ASST.PROFESSOR,
DADI INSTITUTE OF ENGG.&TECH.,
ANAKAPALLI,
VISAKHAPATNAMDISTRICT-531002.