



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Aggregated Approach of Extracting Evidences to Detect Ranking Fraud

Anamika R. Ghatge¹, Prof. U. A. Mande²

P.G. Student, Department of Computer Engineering, Sinhgad College of Engineering, Vadgaon(BK), Pune, India¹

Associate Professor, Department of Computer Engineering, Sinhgad College of Engineering, Vadgaon(BK), Pune, India²

ABSTRACT: Nowadays everyone is using smart phones. There is need of various applications to be installed on smart phone. To download application smart phone user has to visit play store such as Google Play Store, Apples store etc. When user visit play store then he is able to see the various application lists. This list is built on the basis of promotion or advertisement. User doesn't have knowledge about the application (i.e. which applications are useful or useless). So user looks at the list and downloads the applications. But sometimes it happens that the downloaded application won't work or not useful. That means it is fraud in mobile application list. To avoid this fraud, we are making application in which we are going to list the applications. To list the application first we are going to find the active period of the application named as leading session. We are also investing the three types of evidences: Ranking based evidence, Rating based evidence and Review based evidence. Using these three evidences finally we are calculating aggregation. We evaluate our application with real world data collected form play store for 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 at a breath taking 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 daily App leader boards, which demonstrate the chart rankings of most popular Apps. Indeed, the App leader board is one of the most important ways for promoting mobile Apps. A higher rank on the leader board usually leads to a huge number of downloads and million dollars in revenue. Therefore, App developers tend to explore various way such as advertising campaigns to promote their Apps in order to have their Apps ranked as high as possible in such App leader boards. 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 "boot farms" 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 could be propelled from number 1,800 to the top 25 in Apple's top free leader board and more than 50,000-100,000 new users could be acquired within a couple of days. In fact, such ranking fraud raises great concerns to the mobile App industry. For example, Apple has warned of cracking down on App developers who commit ranking fraud in the Apple's App store.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 4, Issue 12, December 2016

II. LITERATURE SURVEY

1) *Paper Name:-Detecting Product Review Spammers using RatingBehaviors*

Authors:-Ee-Peng Lim, Viet-An Nguyen,Nitin Jindal, Bing Liu, Hady W. Lauw

Abstract:-

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 sub-set 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 spammers and 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.

2) *Paper Name:-A taxi driving fraud detection system*

Authors:- Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou

Abstract:-

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.

From this paper we have referred:-

- Concept of fraud detection

3) *Paper Name:- Opinion spam and analysis*

Authors:- N. Jindal and B. Liu

Abstract:-

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



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 4, Issue 12, December 2016

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.

4) *Paper Name:-An unsupervised learning algorithm for rank aggregation*

Authors: -A. Klementiev, D. Roth, and K. Small

Abstract:-

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.

III. EXISTING SYSTEM APPROACH

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. Generally speaking, the related works of this study can be 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. For example, we have studied various aspects of content based spam on the web and presented a number of heuristic for detecting content based spam.

Disadvantages:-

- 1) Occurs Fraud in Web Application.
- 2) Leader Board gets false rating reviews of user.

IV. PROPOSED SYSTEM APPROACH

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 4, Issue 12, December 2016

ADVANTAGES OF PROPOSED SYSTEM

1. The proposed framework is scalable and can be extended with other domain generated evidences for ranking fraud detection.
2. Experimental results show the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity of ranking fraud activities.
3. 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).

V. SYSTEM ARCHITECTURE

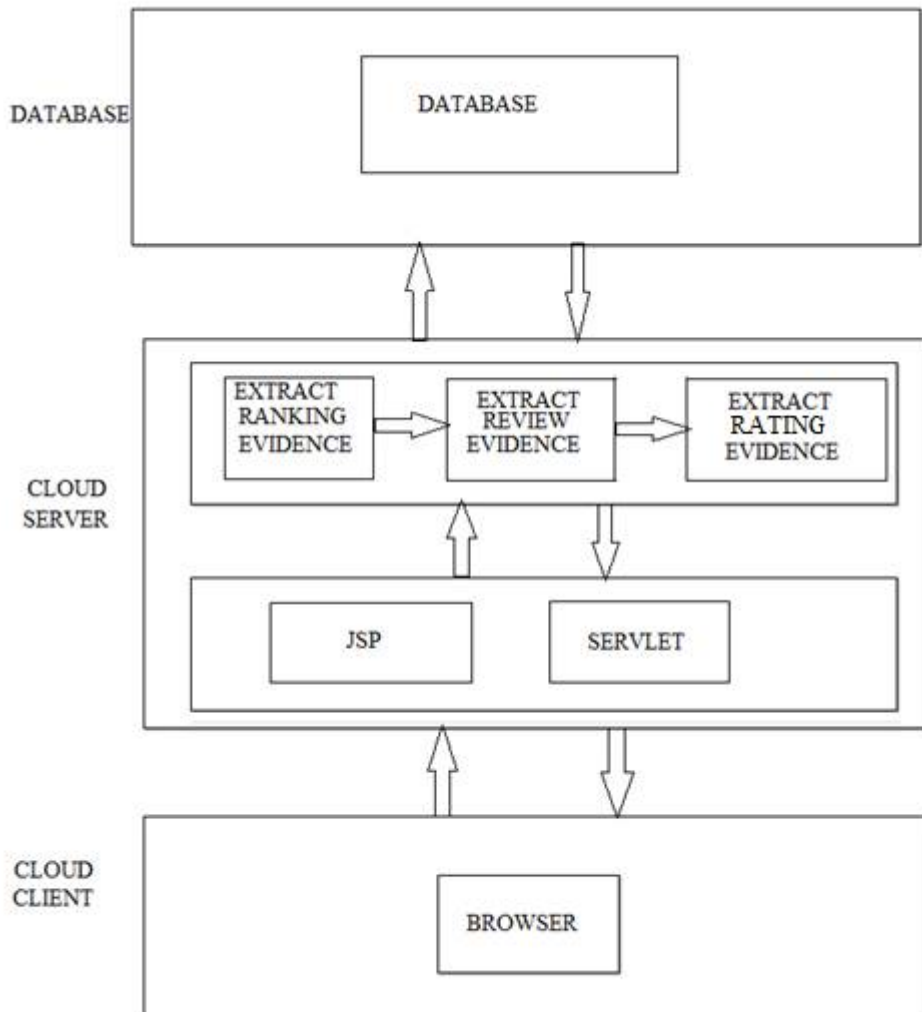


Fig No 01 System Architecture



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 4, Issue 12, December 2016

VI. CONCLUSION

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. A unique perspective of this approach is that all the evidences can be modeled by statistical hypothesis tests, thus it is easy to be extended with other evidences from domain knowledge to detect ranking fraud. Finally, we validate the proposed system with extensive experiments on real-world App data collected from the Google Play store.

REFERENCES

- [1] Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, Hady W. Lauw, "Detecting Product Review Spammers using Rating Behaviors", CIKM'10, October 26–30, 2010.
- [2] 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.
- [3] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219–230.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] G. Shafer, A Mathematical Theory of Evidence. Princeton, NJ, USA: Princeton Univ. Press, 1976.
- [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] 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.