DISCOVERY OF RAKING FRAUD DETECTION FOR MOBILE APPLICATION

[1]Prof. AnikitSanghvi, [2]Suraj Singh, [3]AbhijeetKonkar, [4]AshishGavit
[1]HOD. of Computer Engg. Dept., [2,3,4]Student of IT Engg.Dept.
AlamuriRatnamala Institute of Engineering and Technology, Asangaon, Shahapur

Abstract: the mobile apps growth is increasing day by day and it's a well know concept now days. Due to huge number of mobile apps in the market, finding ranking fraud is a big task in front of the mobile Apps market. Ranking fraud means fake or unwanted activity which have intention to up the Apps in the popularity list. In the existing system the leading event and leading session of an app is recognize from the collected past records. Based on the users feedback, tree different types of evidence taking into consideration, 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 Approval of scores by the admin to identify the exact reviews and rating 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 IP address that limits the number of user login logged per day. 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

There are many Apps in the Apps market and there are some apps which are non-genuine. Due to statistics of Google play store , there are 3.5 million Apps on Google play store till year 2017. There are huge possibility of ranking fraud in mobile application in app market. In this field, study and research about fraud detection are limited. In this paper we proposed the fraud ranking detection in mobile application with the help of three evidences namely ranking based evidence, rating based evidence and review based evidence by modeling apps. We check ranking, rating and review behavior by statistical hypotheses tests. In additional, we proposed an optimization based aggregation method to combine all the evidences for fraud detection. With the help of this application we can easily identify the fraud application in the apps market and save our time, data, and money (if paid). 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, rating and review history, which gives some anomaly patterns fromapps historical rating and reviews records.

II. LITERATURE REVIEW

In recent years the quantity of versatile Apps has developed at a drastically fast rate. Due to the rapid advancement in the mobile technology and mobile devices, mobile App is a very popular and well known concept. Due to the popularity, mobiles are major target for malicious applications. Main challenges in the popular operating system called android are to detect and remove malicious apps. Many App stores launched daily App leaderboard to rank most popular Apps which inspired the development of mobile Apps. Apps which are on the top list of the leader board in turn lead to a large number of downloads and million dollars in profits. Thus, App developers computing for various ways like advertising drive to support their Apps in order to get their Apps ranked as high as possible in such App leaderboards. Usually in the market dishonest App developers for Apps enhancement fraudulent means to consciously boost their apps and distort the chart rankings on an App store. To implement this novel solution provided called bot forms and human water armies to increase the App downloads, ratings and reviews in a very short time.

A. Novel technique for computing a rank aggregation

Rank aggregation via nuclear norm minimization in this paper author D. F. Gleich and L.-h. Lim.proposed novel technique for computing a rank aggregation on the basis of matrix completion to avoid noise and incomplete data. Proposed method solves a structured matrix-completionproblem overthe space of skew-symmetric matrices. Author proves a recovery theorem detailing when proposed approach will work. They also perform a detailed evaluation of proposed approach with synthetic data and an anecdotal study with Netflix ratings. To find the solutions, they utilized the svp solver for matrix completion. Rank aggregation is combined with structure of skew-symmetric matrices. Author applied latest advances in the theory and algorithms of matrix completion to skew-symmetric matrices. Author enhanced existing algorithm for matrix completion to handle skew-symmetric data[2].

B. Unsupervised learning algorithm

The author A. Klementiev, D. Roth, and K. Small proposed a novel method to the rank aggregation problem by providing an optimization issues to discover a linear combination of ranking functions which exploits agreement. To solve this problem author introduce an unsupervised learning algorithm called ULARA which returns a linear combination of the individual ranking functions based on the principle of rewarding ordering agreement between the rankers. Effectiveness of the proposed technique is measure based on a data fusion task across ad hoc retrieval systems[3].

C. Aggregating rankings without supervision

The A. Klementiev, D. Roth author proposed novel a formal mathematical and algorithmic framework to with purpose of aggregating rankings without supervision. Various key challenges in heuristic and supervised learning approaches are presents as they require domain knowledge or supervised ranked data. Author designed an EM-based algorithm and illustrates that it can be made efficient for the right-invariant decomposable distance functions. Proposed scheme is effective than the other existing scheme. For efficient learning author design the concept of augmented permutation and a novel decomposable distance function. Proposed framework is applicable to other types of partial rankings, as well as to situations where ranking data is not of the same type[4].

D. Review spammers

The author E.-P. Lim, V.-A. Nguyen proposed a novel scheme to detect review spammers who try to influence review ratings on some target products or product groups. Main objective of the system is to detect users generating spam reviews or review spammers. Author analyses the features behaviours of review spammers and on the basis of that they utilize this behaviour to detect the spammers. Author model the certain behaviours such as first, spammers may target specific products or product groups for maximizing their impact, secondly they tend to deviate from the other reviewers in their ratings of products. Author proposed the scoring methods to compute the degree of spam for each reviewer and relate them on an Amazon review dataset. After evaluation it proves that proposed ranking and supervised methods are effective in discovering spammers[5]

E. Bayesian framework

The author A. Mukherjee, A. Kumar proposed the to detect opinion spammers in an unsupervised Bayesian inference framework, author proposed a novel and principled method called Author Spamicity Model (ASM). Proposed scheme is novel as existing methods are mainly depends upon heuristics or ad-hoc labels to detect opinion spam. The Bayesian framework makes possible characterization of many behavioral phenomena of opinion spammers by utilizing the estimated latent population distributions. Solution provided the author is not yet done by any of the existing methods. The results across both evaluation metrics show that the proposed model is effective and outperforms strong competitors[6].

F. Semi-supervised learning system

To detect hybrid shilling attack detection author Z.Wu, J.Wu, J. Cao proposed novel scheme called HySAD. As proposed scheme is of type a semi-supervised learning system which facilitates both unlabeled and labeled user profiles for multi-class modeling. Main benefits provided the proposed scheme is that it provides effective solution against hybrid attacks even though it presented with obfuscated strategies. Author also compares effectiveness with realistic case study on ‘Amazon.cn’ with technology like HySAD in improving the performance of a collaborative-filtering recommender system, and the ability of HySAD to help explore interesting attacker behaviours [7]

Sr. no.	Paper name	Proposed	Advantages	Basic Method used
1.	Rank aggregation via nuclear norm minimization. [2]	author proposed novel technique for computing a rank aggregation on the basis of matrix completion to avoid noise and incomplete data	Proposed method solves a structured matrix-completion problem over the space of skew-symmetric matrices.	Matrix operations
2.	An unsupervised learning algorithm for rank aggregation [3]	Author proposed a novel method to the rank aggregation problem by providing an optimization issues to discover a linear combination of ranking functions which exploits agreement.	Effectiveness of the proposed technique is measure based on a data fusion task across ad hoc retrieval systems.	Linear Combination of Ranking functions
3.	Unsupervised rank aggregation with distance-based	Author proposed novel a formal mathematical and algorithmic framework to	Proposed scheme is effective than the other existing scheme.	Augmented permutation and a novel decomposable

	models [4]	with purpose of aggregating rankings without supervision		distance function
4.	Detecting product review spammers using rating behaviors [5]	Author proposed a novel scheme to detect review spammers who try to influence review ratings on some target products or product groups.	After evaluation it proves that proposed ranking and supervised methods are effective in discovering spammers.	The scoring methods to compute the degree of spam for each reviewer.

TABLE: COMPARISION OF LITERATURE

III.PROPOSED SYSTEM

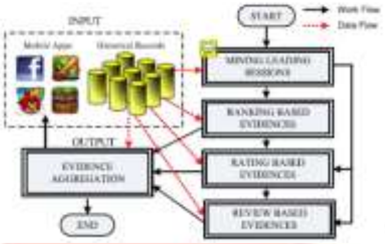
- We first propose a simple yet effective algorithm to identify the leading sessions of each App based on its historical ranking records. Then, with the analysis of Apps' ranking behaviors, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, we characterize some fraud evidences from Apps' historical ranking records, and 0develop 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 perspective of App ranking fraud.



IV. 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. An 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 Apple's App store. Experimental results showed the effectiveness of the proposed approach. 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.

REFERENCES

[1] Hengshu Zhu, HuiXiong, Senior Member, IEEE, Yong Ge, and Enhong Chen, Senior Member, IEEE,"Discovery of

Ranking Fraud for Mobile Apps", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO.1, January 2015.Learn. Res. 3 (Mar. 2003), 1289-1305.

[2] D. F. Gleich and L.-h. Lim. Rank aggregation via nuclear norm minimization. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '11, pages 60–68, 2011.Y.T. Yu, M.F. Lau, "A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions", Journal of Systems and Software, 2005, in press.

[3] A. Klementiev, D. Roth, and K. Small.An unsupervised learning algorithm for rank aggregation. In Proceedings of the 18th European conference on Machine Learning, ECML '07, pages 616–623, 2007.

[4] A. Klementiev, D. Roth, and K. Small.Unsupervised rank aggregation with distance-based models. In Proceedings of the 25th international conference on Machine learning, ICML '08, pages 472– 479, 2008.

[5] 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.

[6] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh.Spotting opinion spammers using behavioral footprints.In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '13, 2013.

[7] 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.S