

Survey on Fraud Ranking Detection in mobile app store

M. Aadil Khan
SKN College of Engineering
Vadgaon, Pune-410507

T.H. Gurav
SKN College of Engineering
Vadgaon, Pune-410507

ABSTRACT

From last few years, mobile technology has been received much more attention since it is most popular and basic need of today's world. Due to the popularity, mobiles are major target for malicious applications. Key challenge is to detect and remove malicious apps from mobiles. Numerous amounts of mobile apps are generated daily so ranking fraud is the one of the major aspects in front of the mobile App market. Ranking fraud refers to fraudulent or vulnerable activities. Main aim of the fraudulent is to knock the fraud mobile apps in the popularity list. Most App developer generates the ranking fraud apps by tricky means like enhancing the apps sales or by simply rating fake apps. Thus, there is need to have novel system to effectively analyze fraud apps.

This paper provides a survey on various existing techniques with the novelties highlighting the need of novel technique to detect fraud mobile apps. This paper is motivated by arising need to detect fraud apps with less time. In proposed system, we add recommendation based on the modified ranking.

General Terms

Data Mining, Information Security, Hypothesis and their p-values.

Keywords

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

1. INTRODUCTION

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.

Huge amount of mobile apps are generated daily so ranking fraud is the one of the major aspects in front of the mobile App market. So it is today need to develop the novel solution to detect fraud mobile Apps. In section II there is a study of existing methods and their literature survey. In chapter III a brief discussion is done on our detection system with

recommendations. At last in section IV there is a conclusion given for this survey.

2. LITERATURE SURVEY

In paper [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. 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.

In paper [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. 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.

Proposed algorithm effectiveness is measured using the two experimental settings that utilized two functions such as on synthetic data which compute performance with Spearman's rank correlation coefficient and an information recovery data fusion task which quantifies performance by utilizing precision.

In paper [4], 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.

In paper [5], author 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 analyzes the features behaviors of review spammers and on the basis of that they utilize this behavior to detect the

spammers. Author model the certain behaviors 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.

In paper [6], 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.

In paper [7], to detect hybrid shilling attack detection author 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 behaviors.

In paper [8], author studied the problem of unsupervised web spam detection. Author initiates the notion of spamicity to compute how likely a page is spam. Author proposed efficient link spam and term spam detection methods and it does not require training and are cost effective. After evaluation with the real data set it is clear that proposed methods are effective and efficient to detect spam pages.

Table 1: Comparison of literature

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 models [4]	Author proposed novel a formal mathematical and algorithmic framework to with purpose of aggregating rankings without supervision	Proposed scheme is effective than the other existing scheme.	Augmented permutation and a novel decomposable 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.
5.	Spotting opinion spammers using behavioral footprints [6]	To detect opinion spammers in an unsupervised Bayesian inference framework, author proposed a novel and principled method called Author Spamicity Model (ASM).	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	Heuristics or ad-hoc label based unsupervised Bayesian inference framework

3. EVIDANCE AGGRIGATION

The study describes a ranking fraud detection process where there are some evidences considered and integrated to get an

aggregated result which is most reliable in finding a fraudulent application in a mobile market[1]. Most generally the ranking fraud is happening in some particular phase of an application, it is called as a leading session comprised of

many leading events[1]. A leading event may occur due to an advertisement campaign or etc.

This study can be extended to get a recommender system to enhance user experience.

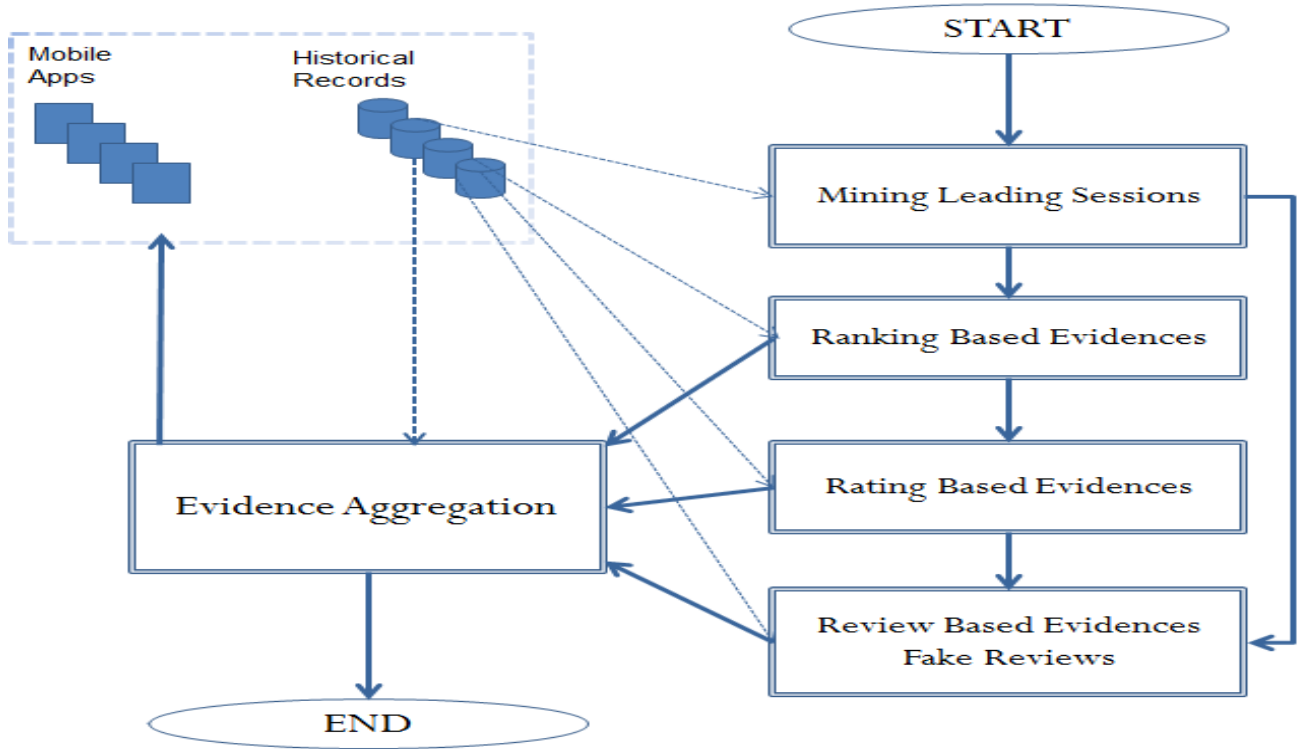


Fig 1: System architecture

4. PROPOSED SYSTEM

Ranking fraud detection task is a very valuable work in this area. The existing technique integrate the results of various evidences to find genuineness of an app. We again integrate a recommender system and the ranking score generated

previously such that, the recommender system will recommend the most genuine app that are most relevant. The apps suggested by recommender system will be checked for whose score is high, and the app with highest score will be suggested.

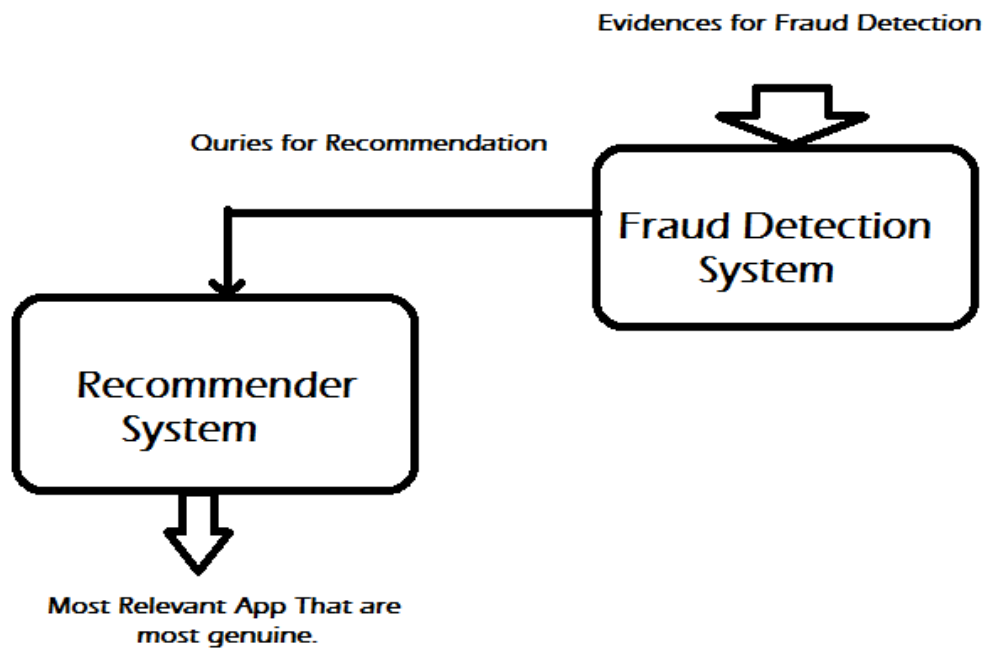


Fig 2: Proposed architecture

5. CONCLUSION AND FUTURE SCOPE

This paper presents an all-inclusive survey on fraud detection in mobile Apps. The main features, the advantages and disadvantages of each system are described. In recent years the quantity of versatile Apps has developed at a breath taking rate. Business and research communities are significantly attracted by this fake review detection. As increase popularity, mobiles are major target for malicious applications. Main challenge is to detect and remove malicious apps from mobile app market. Main aim of the fraudulent is to knock the fraud mobile apps in the popularity list.

Thus, there is need to have novel system to effectively analyze fraud apps. In proposed system, system performance can be enhanced by adding the recommendation based on the ranking.

In Future a study can also be done to make a detection system that tracks the online campaigning on social media. Online campaigning may be used to gain a particular benefit that may be of business, politics or something else.

6. REFERENCES

- [1] Hengshu Zhu, Hui Xiong, 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
- [8] 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.
- [9] H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian. Exploiting enriched contextual information for mobile app classification. In Proceedings of the 21st ACM international conference on Information and knowledge management, CIKM '12, pages 1617–1621, 2012.
- [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.