RNI Regd No: APENG/2011/47294 ISSN No: 2320-3706(PRINT)
International Journal & Magazine of Engineering, Technology, Management and Research
YUVAENGINEERS
Transforming Young Engineers for Better Tomorrow

Discovery of Ranking Fraud for Mobile Apps

Ch.Ramesh Kumar
Associate Professor & HOD,
Department of CSE,
Malla Reddy Engineering College
& Management Sciences, Kistapur,
Medchal, Hyderabad.

B.Prasanna Jyothi
Assistant Professor,
Department of CSE,
Malla Reddy Engineering College
& Management Sciences, Kistapur,
Medchal, Hyderabad.

Naresh Rosa
M.Tech Student,
Department of CSE,
Malla Reddy Engineering College
& Management Sciences, Kistapur,
Medchal, Hyderabad.

Abstract:
Now days, mobile App is a very popular and well known concept due to the rapid advancement in the mobile technology and mobile devices. Due to the large number of mobile Apps, ranking fraud is the key challenge in front of the mobile App market. Ranking fraud refers to fraudulent or vulnerable activities which have a purpose of bumping up the Apps in the popularity list. In fact, it becomes more and more frequent for App developers to use tricky means, like increasing their Apps' sales or posting fake App ratings, to commit ranking fraud. While the importance and necessity of preventing ranking fraud has been widely recognized.

After understanding the details of ranking fraud and the need of ranking fraud detection, the paper proposes a ranking fraud detection system for mobile Apps. The proposed system mines the active periods such as leading sessions of mobile apps to accurately locate the ranking fraud. These leading sessions can be useful for detecting the local anomaly instead of global anomaly of App rankings. Besides this, by modeling Apps ranking, rating and review behaviours using statistical hypotheses tests, we investigate three types of evidences, they are ranking based evidences, rating based evidences and review based evidences. Furthermore, we propose an aggregation method based on optimization to integrate all the evidences for fraud detection. 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.

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 characterized 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.

LITERATURE SURVEY:

1. A Flexible Generative Model Forpreference Aggregation:

Many areas of study, such as information retrieval, collaborative filtering, and social choice face the preference aggregation problem, in which multiple preferences over objects must be combined into a consensus ranking. Preferences over items can be expressed in a variety of forms, which makes the aggregation problem difficult. In this work we formulate a flexible probabilistic model over pairwise comparisons that can accommodate all these forms. Inference in the model is very fast, making it applicable to problems with hundreds of thousands of preferences. Experiments on benchmark datasets demonstrate superior performance to existing methods.

2.Get Jar Mobile Application Recommendations With Very Sparse Datasets:

The Netflix competition of 2006 [2] has spurred significant activity in the commendations field, particularly in approaches using latent factor models [3, 5, 8, 12] However, the near ubiquity of the Netflix and the similar MovieLens datasets1 may be narrowing the generality of lessons learned in this field. At GetJar, our goal is to make appealing recommendations of mobile applications (apps). For app usage, we observe a distribution that has higher kurtosis (heavier head and longer tail) than that for the aforementioned movie datasets. This happens primarily because of the large disparity in resources available to app developers and the low cost of app publication relative to movies.

In this paper we compare a latent factor (Pure SVD) and a memory-based model with our novel PCA-based model, which we call Eigen app. We use both accuracy and variety as evaluation metrics. Pure SVD did not perform well due to its reliance on explicit feedback such as ratings, which we do not have. Memory-based approaches that perform vector operations in the original high dimensional space over-predict popular apps because they fail to capture the neighborhood .

3. Detectingspam Web Pages Through Content Analysis:

In this paper, we continue our 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-un described techniques for automatically detecting spam pages, examines the effectiveness of these techniques in isolation and when aggregated using classification algorithms. When combined, our heuristics correctly identify 2,037 (86.2%) of the 2,364 spam pages (13.8%) in our judged collection of 17,168 pages, while misidentifying 526 spam and non-spam pages (3.1%).

4.Spotting Opinion Spammers Using Behavioral Footprints:

Opinionated social media such as product reviews are now widely used by individuals and organizations for their decision making. However, due to the reason of profit or fame, people try to game the system by opinion spamming (e.g., writing fake reviews) to promote or to demote some target products. In recent years, fake review detection has attracted significant attention from both the business and research communities. However, due to the difficulty of human labeling needed for supervised learning and evaluation, the problem remains to be highly challenging.

This work proposes a novel angle to the problem by modeling spamicity as latent. An unsupervised model, called Author Spamicity Model (ASM), is proposed. It works in the Bayesian setting, which facilitates modeling spamicity of authors as latent and allows us to exploit various observed behavioral footprints of reviewers. The intuition is that opinion spammers have different behavioral distributions than non-spammers. This creates a distributional divergence between the latent population distributions of two clusters: spammers and non-spammers. Model inference results in learning the population distributions of the two clusters. Several extensions of ASM are also considered leveraging from different priors. Experiments on a real-life Amazon review dataset demonstrate the effectiveness of the proposed models which significantly outperform the state-of-the-art competitors.

5.Unsupervisedrank Aggregation With Domain-Specific Expertise:

Consider the setting where a panel of judges is repeatedly asked to (partially) rank sets of objects according to given criteria, and assume that the judges' expertise depends on the objects' domain. Learning to aggregate their rankings with the goal of producing a better joint ranking is a fundamental problem in many areas of Information Retrieval and Natural Language Processing, amongst others. However, supervised ranking data is generally difficult to obtain, especially if coming from multiple domains. Therefore, we propose a framework for learning to aggregate votes of constituent rankers with domain specific expertise without supervision. We apply the learning framework to the settings of aggregating full rankings and aggregating top-k lists, demonstrating significant improvements over a domain-agnostic baseline in both cases.

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.

Limitatios:

•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.

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 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 leader board. 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.

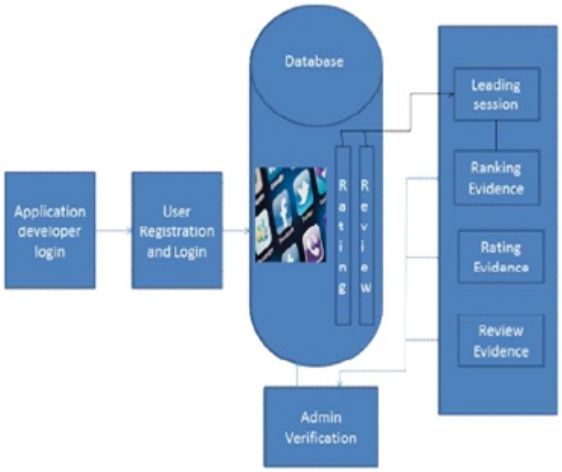


Fig: Architecture diagram

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).

IMPLEMENTATION

1)Mining Leading Sessions:

•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.

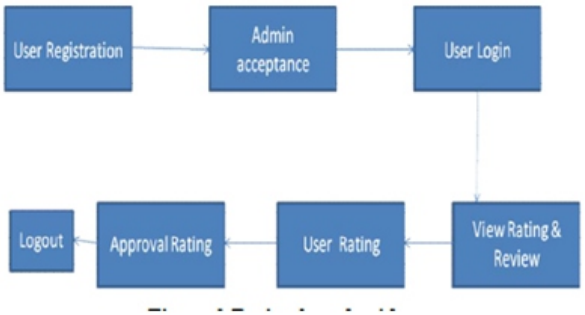
•Therefore, the problem of detecting ranking fraud is to detect fraudulent leading sessions. Along this line, the first task is how to mine the leading sessions of a mobile App from its historical ranking records.
•There are two main steps for mining leading sessions. First, we need to discover leading events from the App's historical ranking records. Second, we need to merge adjacent leading events for constructing leading sessions.

2)Ranking Based Evidences:

•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 leader board (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).

3)Rating Based Evidences:

•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.



4)Review Based Evidences:

•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 an often post fake review in the leading sessions of a specific App in order to inflate the App downloads, and thus propels the App's ranking position in the leader board.

5) Evidence Aggregation:

•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.

CONCLUSION:

This paper reviews various existing methods used for web spam detection, which is related to the ranking fraud for mobile Apps. Also, we have seen references for online review spam detection and mobile App recommendation. By mining the leading sessions of mobile Apps, we aim to locate the ranking fraud. The leading sessions works for detecting the local anomaly of App rankings. The system aims to detect the ranking frauds based on three types of evidences, such as ranking based evidences, rating based evidences and review based evidences. Further, an optimization based aggregation method combines all the three evidences to detect the fraud.

REFERENCES:

[1] (2012). [Online]. Available: http://venturebeat.com/2012/07/03/ apples-crackdown-on-app-ranking-manipulation/ .

[2] (2012). [Online]. Available: https://developer.apple.com/news/ index.php?id=02062012a.

[3] 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.

[4] 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.

[5] B. Zhou, J. Pei, and Z. Tang, "A spamicity approach to web spam detection," in Proc. SIAM Int. Conf. Data Mining, 2008, pp. 277– 288.

[6] 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.

[7] Z. Wu, J. Wu, J. Cao, and D. Tao, "HySAD: A semi-supervised hybrid shilling attack detector for trustworthy product recommendation," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 985–993.

[8] 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.

[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] B. Yan and G. Chen, "AppJoy: Personalized mobile application discovery," in Proc. 9th Int. Conf. Mobile Syst., Appl., Serv., 2011, pp. 113–126.

[11] H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian, "Exploiting enriched contextual information for mobile app classification," in Proc. 21st ACMInt. Conf. Inform. Knowl. Manage., 2012, pp. 1617–1621.

[12] 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.

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

Authors Biography:

Naresh Rosa, Completed his B.Tech degree in Pulla Hasvita Institute of Engineering and Technology in 2014. He is pursuing M.Tech. in Computer Science & Engineering from Department of Computer Science & Engineering in Malla Reddy Engineering College & Management sciences, Kistapur, Medchal, Hyderabad. Affiliated to JNTUH, HYDERABAD, TELANGANA, India.. His research interest include cloud, data mining, Big Data and networking.

B.Prasanna Jyothi, Working as Assistant Professor, department of computer science and engineering, in Malla Reddy Engineering College & Management Sciences, Kistapur, Medchal ,Hyderabad. Affiliated to JNTUH, HYDERABAD, TELANGANA., India. Her research interests include data mining, computer networks.

Ch.Ramesh Kumar, Working as Assoc. Prof & Head of the Department of Computer Science and Engineering in Malla Reddy Engineering College & Management Sciences, Kistapur, Medchal, Hyderabad. Affiliated to JNTUH, HYDERABAD, TELANGANA., India. he has several international publications to his credit. His research interests include Software reuse, Software performance, Software testing ,Data Mining and cloud computing.