ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A New Approach for Mobile App with Discovery of Ranking Fraud Detection

M.Dharma Vardhani
Lecturer,
Dept of Computer Science,
Sri Durga Malleswara Siddhartha Mahila Kalasala,
Vijayawada, A.P., India.

K.Parish Venkata Kumar, M.Tech, (Ph.D)
Assistant Professor,
Department Of Computer Applications,
V.R.Siddhartha Engineering College,
Kanuru,Vijayawada-520 007, A.P., India.

Abstract:
Smart phone users has to download application, to visit Apps store such as Google Play Store, Apples store etc. When user visit play store then he or she is able to see the various applications list. This list is built on the basis of promotion or advertisement. Ranking fraud in the mobile Apps is nothing but the false or fake activities which are going to be done in Apps popularity list for bumping up the fraud App . It is very easy for the App developer to use the fake App and fake rating for committing the ranking fraud . This paper gives holistic view of ranking manipulations and portrays a Ranking fraud identifiable framework for mobile Apps. This work is done into three classifications. At first, ranking fraud discovery second is online review identification and last one is mobile application suggestions.

Proposed framework additionally eliminates the fake surveys from the dataset utilizing same measure algorithm and after that distinguish the application rank. At last this system will also recommend Apps which are more relevant and most genuine. The propose framework will saves the time and also memory than the previous framework. we evaluate the proposed system with real-world App data collected from the Google App Store for a long time period. There are mainly three types of evidences, Ranking based evidences , Rating based evidences , and Review based evidences . These evidences can be obtained from Apps ranking , rating and review history .Then proposed an optimization based aggregation method for integrating all these evidences for the fraud detection.

In this way more effectiveness and regularity of the ranking fraud detection system is obtained and the original App is recommended to the user.

Keywords:
Mobile Apps, Ranking Fraud Detection, Review and Rating.

I. INTRODUCTION:

Many mobile app stores launched day by day app leader boards which shows the chart ranking of popular apps. The leader board is the important for promoting apps. Original application grade level decreases due to the duplication arrival in the mobile apps. In recent activities duplicate version of an application not burned or blocked. This is the major defect. Higher rank leads huge number of downloads and the app developer will get more profit. In this way they allow Fake Application also. User not understanding the Fake Apps then the user also give the reviews in the fake application. Exact Review or Ratings or Ranking Percentage are not correctly Calculated. Positioning coercion in the convenient App business division insinuates misleading or beguiling practices which have an inspiration thumping up Apps in the notoriety list. Without a doubt, it turns out to be more continuous for App developers to utilize shady. Smart phones emerges new technologies like android and iOS operating system took a boost in market. Mobile application started growing at such a high rate. As a study says millions of apps are there on apple's app store and on Google Play. This started a new business in computer world and became a reason to earn thousands of dollars and downloads.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Daily leader board is published by these markets contains the most popular apps which will consequently be downloaded and rated most high by users. Some developers may use some marketing strategies like an advertisement campaign for promotion of their app. However this part of technology is also not safe from threats. Mobile app market, we refer it as market, is manipulated by some fraudulent app developers to bump up their app high in the rank list, as an app in leader board confirms high downloads and high income. Shady means are used to make such a fraud and implemented using "bot farms" which is also called "Human water armies". In this area some related work is there, for example, spam detection for web ranking, mobile app recommendations, and some online review based spam detection. Our study thus focuses on an integrated approach, for various evidences, to find Mobile App ranking fraud and also recommend the most relevant App that is most genuine.

For this we have to go through challenges like first we need to find at what time the fraud is happening it means exact time of fraud is needed. Secondly we know that there are tremendous number of Apps present in market so it is nearly impossible to physically mark ranking fraud for every App, so it's crucial to automatically distinguish fraud without utilizing any essential data. 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. To this end, 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 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. Especially, 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 from apps historical rating and reviews records.

II . PROPOSED SYSTEM:

The main target is to detect ranking fraud for the mobile App.Mainly fraud happens in leading sessions.Thats why rst propose the effective algorithm for identifying leading sessions of App which is truly based on the ranking history . after that Apps ranking behaviours is analyzed for nding the fake App . Because the fake Apps have different ranking records as compared with the normal App . Therefore from Apps history of ranking records we have to characterize some fraud evidences and then develop different functions for extracting ranking based fraud evidences .There are two types in fraud evidences which based on Apps rating and the review history . In proposed system we overcome the drawbacks of Mining leading session algorithm which is based on ranking, review & rating. Detection of ranking fraud for mobile Apps is still under a subject to research.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

To fill this crucial lack, we propose to develop a ranking fraud detection system for mobile Apps. We also determine several important challenges. First challenge, in the whole life cycle of an App, the ranking fraud does not always happen, so we need to detect the time when fraud happens. This challenge can be considered as detecting the local anomaly in place of global anomaly of mobile Apps. Second challenge, it is important to have a scalable way to positively detect ranking fraud without using any basis information, as there are huge number of mobile Apps, it is very difficult to manually label ranking fraud for each App. Finally, due to the dynamic nature of chart rankings, it is difficult to find and verify the evidences associated with ranking fraud, which motivates us to discover some implicit fraud patterns of mobile Apps as evidences.

III .RELATED WORK:

1. Ranking Based Evidences:

A leading session is composed of several leading events. Therefore, we should first analyze the basic characteristics of leading events for extracting fraud 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. 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).

2. Rating Based Evidences:

The ranking based evidences are useful for ranking fraud detection. However, sometimes, it is not sufficient to only use ranking 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. Intuitively, if an App has ranking fraud in a leading session s, the ratings during the time period of s may have anomaly patterns compared with its historical ratings, which can be used for constructing rating based evidences.

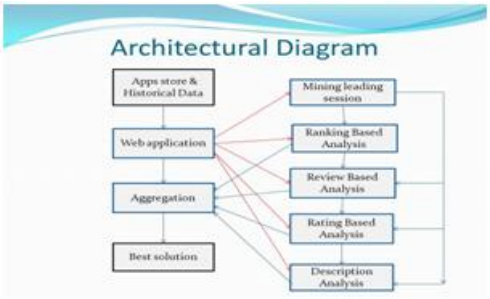
3. Review Based Evidences:

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. Specifically, before downloading or purchasing a new mobile App, users often firstly 5, 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 downloads, and thus propel the App's ranking position in the leader board. Although some previous works on review spam detection have been reported in recent years, the problem of detecting the local anomaly of reviews in the leading sessions and capturing them as evidences for ranking fraud detection are still under-explored.

4. Identifying the leading sessions for mobile apps:

Basically, mining leading sessions has two types of steps concerning with mobile fraud apps. Firstly, from the Apps historical ranking records, discovery of leading events is done and then secondly merging of adjacent leading events is done which appeared for constructing leading sessions. Certainly, some specific algorithm is demonstrated from the pseudo code of mining sessions of given mobile App and that algorithm is able to identify the certain leading events and sessions by scanning historical records one by one.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal



IV. LITERATURE SURVEY:

1) A flexible generative model for preference Aggregation

AUTHORS: M. N. Volkovs and R. S. Zemel

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) Getjar mobile application recommendations with very sparse datasets

AUTHORS: K. Shi and K. Ali

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 (PureSVD) and a memory-based model with our novel PCA-based model, which we call Eigenapp. We use both accuracy and variety as evaluation metrics. PureSVD 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 of less popular apps. They have high accuracy due to the concentration of mass in the head, but did poorly in terms of variety of apps exposed. Eigenapp, which exploits neighborhood information in low dimensional spaces, did well both on precision and variety, underscoring the importance of dimensionality reduction to form quality neighborhoods in high kurtosis distributions.

3) Detecting spam web pages through content analysis

AUTHORS: A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly

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-undescribed 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

AUTHORS: A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos,and R. Ghosh

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

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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.

V.CONCLUSION &FUTURE WORK:

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, Hui Xiong,Yong Ge, and Enhong Chen, "Discovery of Ranking Fraud for Mobile Apps"in Proc. IEEE 27th Int. Conf. Transactions on knowledge and data engineering, 2015, pp. 74-87.

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

[3] 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.

[4] 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.

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

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

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.