ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2016**

# Investigation of Ranking, Rating and Review for Ranking Mobile App Fraud Detection

Vishakha Kothari, Prof. Prabhudev Irabashetti,

ME, Dept. of Computer Engineering, VACOE, Ahmednagar, Maharashtra, India

Asst. Professor, Dept. of Computer Engineering, VACOE, Ahmednagar, Maharashtra, India

**ABSTRACT**: Ranking extortion in the portable App business sector alludes to fake or misleading exercises which have a motivation behind knocking up the Apps in the fame list. For sure, it turns out to be more successive for App designers to utilize shady means, for example, swelling their Apps' business or posting fake App appraisals, to submit positioning extortion. While the significance of averting positioning extortion has been broadly perceived, there is restricted comprehension and examination here.

There are so many applications available on internet because of that user can not always get correct or true reviews about the product on internet. The primary aim of this project is to enhance the prevention of ranking frauds in mobile apps using mining and NLP techniques. In the existing system the leading event and leading session of an app is identified from the collected historical records. Then three different types of evidences are collected from the user feedbacks namely ranking based evidence, rating based evidence and review based evidence. These three evidences are aggregated by using evidence aggregation method.

**KEYWORDS**: Mobile Apps, ranking fraud detection, evidence aggregation, historical ranking records, rating and review.

## I. INTRODUCTION

Now a day, most of us uses android Mobile and also uses the play store capability normally. Play store provide great number of application but unluckily few of those applications are fraud. Such applications damage to phone and also may be data thefts. Hence such applications must be marked, so that they will be identifiable for play store users. Ranking extortion in the portable App business sector alludes to fake or misleading exercises which have a motivation behind knocking up the Apps in the renowned list.

So we are proposing an application which will process ranking based evidence, rating based evidence and review based evidence with stop words removal, NLP and mining techniques. So it will be easier to decide which application is fraud or not. There are more than 1.6 million Apps at Apple's App store and Google Play.

Instead of depending on traditional marketing solutions, shady App developers resort to some fraudulent means to intentionally boost up their Apps and ultimately influence the chart rankings on an App store. This is executed by using Bot-farms or human water armies to increase the App downloads, ratings and reviews in a very short time.

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.

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.

- Google Play Store Parser:

Take out the information from Google play store. Provide the URL path of apps, single or multiple path. Extracts following property of android application

- Ranking
- App Category

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2016**

- App Version
- Size
- Installs
- Review
- Rating

We are proposing system with 2 enhancements. Firstly, we are using Senti-word dictionary to identify the exact reviews 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 limited number of user login logged per day into system.

So we have 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.

## 1. Literature Review: [Paper 1]

A mobile developer intentionally or accidentally keeps the in-app advertising control near to where the user must touch, or scroll on usage of smart phone. With the given micro screen real-estate, the user will be lead to mistap while working on the mobile app.

A Klementiev, D. Roth, and K. Small, produces a model which has to integrate the set of rankings often deals with aggregating and it only comes up when a certain ranked data is developed. Even though the various heuristic and supervised learning approaches to rank aggregation, a requirement of domain knowledge and supervised ranked data exists. Therefore, to solve this issue, a framework is proposed for learning aggregate rankings without supervision. This framework is instantiated for the cases of permutations and combinations of top-k lists [4].

D. M. Blei, A. Y. Ng, and M. I. Jordan, introduces a unique model called as Dirichlet allocation (LDA) a generative probabilistic model for collections of discrete data such as text amount. Basically it is a three level hierarchical Bayesian model in which each element of a group is demonstrated as a finite mixture over a fundamental set of topics. Each topic is demonstrated as an infinite mixture over fundamental set of topic probabilities. With the reference of text modeling, the topic probabilities provide an open representation of a document. An efficient approximation inference technique is presented based on various methods and an EM algorithm for empirical Bayes parameter estimation is also presented. The results are reported in document modeling, text classification and collaborative filtering, which compares to a collection of unigrams and probabilistic LSI model [6].

Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou , illustrated that growth in the field of GPS tracking technology have allowed the users to install GPS tracking devices in taxies to gather huge amount of GPS traces under some time period. These traces by GPS offered an unparalleled opportunity to uncover taxi driving fraud traces and then fraud detection system is proposed which is able to identify taxi driving fraud. First, two sort of function are uncovered here i.e. travel route evidence and driving distance evidence. Even a third function is developed to combine the previous functions based on Dempster-Shafer theory. First identification of interesting locations is done from tremendous amount of taxi GPS logs and then parameter free method is proposed to extract the travel route evidences. Secondly, concept of route mark is +++developed to illustrate the driving path between locations and based on those mark, specific model is characterized for the distribution of driving distance and discover the driving distance evidences. And finally, taxi driving fraud detection system with large scale real world taxi GPS logs [7].

T. L. Griffiths and M. Steyvers, introduces the process of rank aggregation which is interweave with the structure of skew-symmetric matrices. Recent development in the principles of matrix completion matrices is been applied and this idea gives rise to a new method for ranking a set of items. The core of this idea deals with the raking aggregation method which intimately describes a partially filled skew-symmetric matrix. The algorithm for matrix completion is raised to hold skew-symmetric data and use that to extract ranks for each item. This algorithm applies same strategy for both pairwise comparisons as well as for rating data. It becomes robust to noise and incomplete data as it is based on matrix completion [8].

A. Klementiev, D. Roth, and K. Small, describes the field of information retrieval, data mining, and natural language, many applications needs a ranking of instances which is not present in classification. Furthermore, a rank

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2016**

aggregation is a result of aggregating the results of the established ranking models into a formalism and then result represents a novel unsupervised learning algorithm (ULARA)which gives a linear combination of individual ranking functions. These functions were developed based on the axiom of rewarding ordering agreement between the rankers [9].

**Click-spam detection [17] :** Many researchers mainly focused on (on early generation) bots to detect the click-spam. When Sbotminer looked for the anomalies in query distribution, he found search engine bots. Millie and Gillberg reported the unusual collusion in users' associated with different publishers that may be pointing to bot behavior. User-Driven Access Control, Bluff ads and Premium Clicks, aim to authenticate user presence to mitigate click-spam. More general approach is proposed by Vachadev, to target every kind of click-spam including bot and non-bot mechanisms.

Since, 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

Also we can easily conclude from existing system that 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). These 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.

## II. PROPOSED SYSTEM

In recent years, mobile app has been growing tremendously while boosting more than 400,000 applications like Apple app store and Google Android market. This tremendous growth of mobile App has made it difficult to user for finding unique and trusted patterns of Application in crowded App stores. Thus to solve this important issue, existing marketing executives precisely use the App download history and ratings by the users to recommend the mobile applications which is totally trusted. Identifying ranking fraud is actually to identify ranking fraud of mobile apps within such leading sessions.

In this paper, an useful algorithm is used to discover the leading sessions based on the historical records and with the help of analysis of those records, it is proved that deceptive apps usually have different ranking patterns in each leading sessions as compared to the normal apps. Therefore it is illustrated from those ranking records that some fraud is taking place in mobile app market and to restrict those frauds, three main evidences are developed to detect such fraud. As only ranking based evidences does not seems to be much sufficient to detect the fraud of mobile app, based on apps rating and review history some fraud evidences were discovered which showed anomaly patterns by those history. Specifically, an unsupervised evidence aggregation method is also proposed here to integrate all such types of evidences for evaluating the trustworthiness of leading sessions.

And finally, the proposed system is estimated with real world app data gathered from the Apple's app store for time consuming period, i.e., more than two years. The results of these experiments showed an effectiveness of proposed approach in fig 1.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2016**

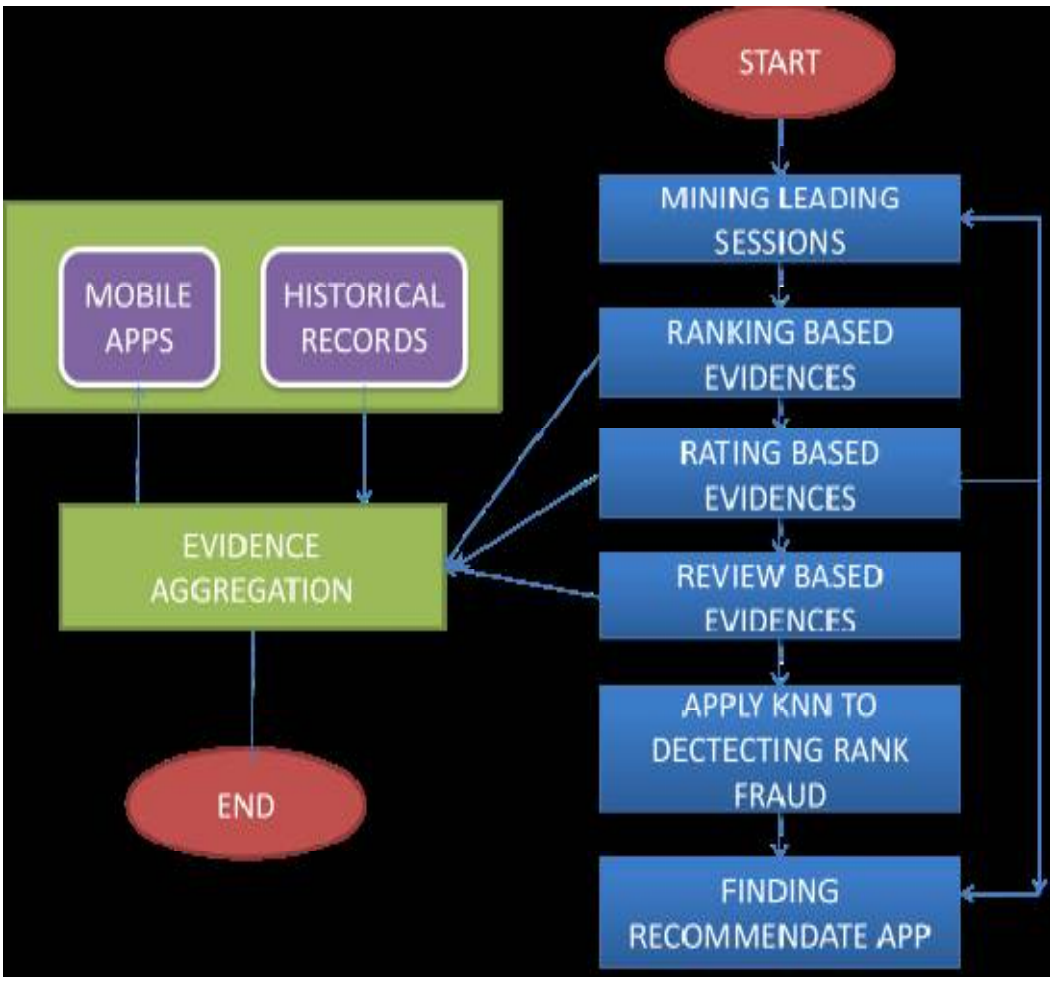


Fig 1: Architecture of Evidence Aggregation based Ranking Fraud Detection.

We are first proposed 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.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Following details block diagram will gives the flow of data.

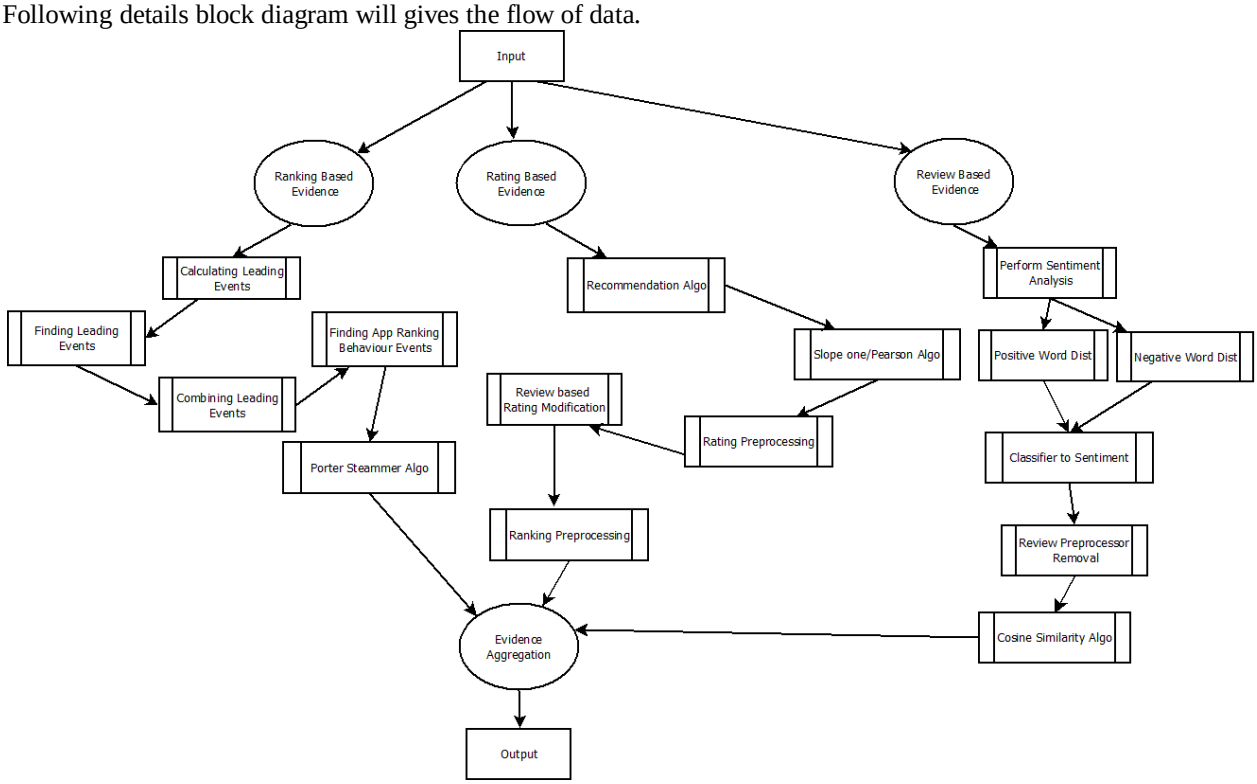


Fig 2: Block diagram of Evidence Aggregation based Ranking Fraud Detection

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.

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. The rating based confirmation is utilized to rate by any client who downloaded it.

The KNN calculation is utilized to enhance effectiveness and precision of the application. These all proofs are consolidated for recognizing the extortion applications. 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.

- **Evidence Aggregation:** 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). Audit based confirmation is utilized to check the surveys of the application.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2016**

We developed an unsupervised evidence-aggregation method to integrate these three types of evidences for evaluating the credibility of leading sessions from mobile Apps. Figure 1 shows the framework of our ranking fraud detection system for mobile Apps.

**2. Advantages of Proposed System:**

- 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.

## III. CONCLUSION

This paper introduces a system which is built up and it is actually a positioning extortion discovery framework for mobile Apps. In particular, initially it is demonstrated that positioning misrepresentation happened in driving sessions and gave a system to digging driving sessions for each App from its chronicled positioning records.

Complaints of an original version of application provider can be undertaken by using Mining Leading Session algorithm. The duplicate version is identified by the admin by means of Historical Records. The admin will also see the date of publication of the apps. When the apps is detected as fraudulently published by the admin then the respective app will be blocked. The user can give the feedback at only once.

Sentiword dictionary is used for finding the exact reviews. The admin can block the fake application. The Review or Rating or Ranking given by users are Correctly Calculated. Hence, a new user who wants to download an app for some purpose can get clear view about the available applications.

## REFERENCES

[1] Discovery of Ranking fraud for mobile apps. Hengshu Zhu,Hui Xiong,Senior members,IEEE,Yong Ge,and Enhong Chen,Senior member,IEEE,IEEE transactions on knowledge and data engineering,vol .27,No.1,January 2015.
[2] Detecting product review spammers using rating behaviors. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. In Proceedings of the 19th ACM international conference on Information and knowledge management.
[3] Supervised rank aggregation. Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li In Proceedings of the 16th international conference onWorld Wide Web.
[4] An unsupervised learning algorithm for rank aggregation, A. Klementiev, D. Roth, and K. Small In Proceedings of the 18th Europeanconference on Machine Learning, ECML '07, pages 616–623, 2007.
[5] An unsupervised learning algorithm for rank aggregation, A. Klementiev, D. Roth, and K. Small In Proceedings of the 18th Europeanconference on Machine Learning, ECML '07, pages 616–623, 2007.
[6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., pp. 993–1022, 2003.
[7] 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.
[8] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Nat. Acad. Sci. USA, vol. 101, pp. 5228–5235, 2004.
[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] Getjar mobile application recommendations with very sparse datasets. K. Shi and K. Ali. In Proceedings of the 18th ACM SIGKDDinternational conference on Knowledge discovery and data mining, KDD '12, pages 204–212, 2012.
[11] Ranking fraud Mining personal context-aware preferences for mobile users. H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian. In Data Mining (ICDM), 2012 IEEE 12th International Conference on, pages1212–1217, 2012.
[12] detection for mobile apps H. Zhu, H. Xiong, Y. Ge, and E. Chen. A holistic view. In Proceedings of the 22nd ACMinternational conference on Information and knowledge management,CIKM '13, 2013.
[13] Exploiting enriched contextual information for mobile app classification, H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian. In Proceedings of the 21st ACM international conference on Information and knowledge management, CIKM '12, pages 1617–1621, 2012.
[14] spammers using behavioral Footprints A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos,and R. Ghosh. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '13, 2013.
[15] Detecting product review spammers using rating behaviors. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw In Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM '10, pages 939–948, 2010.
[16] "Detection of Ranking Fraud for Mobile App and Prevention from User's Recommendation" IJIRCCE, DOI: 10.15680/IJIRCCE.2015. 0309112.
[17] Dr.D.Vasumati, M. Sree Vani, Dr.R.Bhramaramba, and O.Yaswanth Babu "Data Mining Approach to Filter Click-spam in Mobile Ad Networks" ICCDMME'2015 April 20-21, 2015 Bangkok.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 6, June 2016**