International Journal For Technological Research In Engineering
Volume 4, Issue 11, July-2017 *ISSN (Online): 2347 - 4718*

K-MEANS ALGORITHM FOR IDENTIFYING RANKING FRAUD FOR MOBILE APPS

C.Srinivas[1], P.Vijay Kumar[2], CH.Chaithanya[3]
[1]Associate Professor, [2]Assistant Professor, [3]M.Tech Student
Department of CSE, Kakatiya Institute of Technology &Science, Warangal District, Telangana, India.

ABSTRACT: ***Nowadays most are mistreating sensible phone. There's would like of assorted applications to be put in on sensible phone. To transfer application sensible phone user must visit Apps store like Google Play Store, Apples store etc. once user visit play store then he or she is in a position to ascertain the varied applications list. This list is constructed on the idea of promotion or promotion. User doesn't have information regarding the applying (i.e. that applications are helpful or useless). Therefore user appearance at the list and downloads the applications largely from front page of App Store. However generally it happens that the downloaded application won't work or not helpful. Which means its fraud in mobile application list? To avoid this fraud, 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 to list the applying initial we have a tendency to be getting to realize the active amount of the applying named as leading session. We have a tendency to be investment the 3 varieties of proofs: Ranking based mostly evidence, Rating based proof and Review based proof. Also, we propose an aggregation technique 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 together 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

The number of mobile Apps has big at a wide ranging rate over the past few years. As an example, as of the tip of April 2013, there are over 1.6 million Apps at Apple's App store and Google Play. To stimulate the event of mobile Apps, several App stores launched daily App leader boards that demonstrate the chart rankings of preferred Apps. Indeed, the App leader board is one among the foremost necessary ways that for promoting mobile Apps. A better rank on the leader board typically ends up in a large variety of downloads and million greenbacks in revenue. Therefore, App developers tend to explore numerous ways that like advertising campaigns to push their Apps so as to possess their Apps hierarchal as high as double in such App leader boards. However, as a recent trend, rather than looking forward to ancient promoting solutions, shady App developers resort to some deceitful means that to deliberately boost their Apps At last, they also distort the chart rankings on an App store. This can be typically enforced by victimisation alleged "internet bot farms" or "human water armies" to inflate the App downloads and ratings during a very short time. as an example, a writing from Venture Beat [3] reportable that, once associate degree App was promoted with the assistance of ranking manipulation, it can be propelled from no 1,800 to highest twenty five in Apple's top free leader board and over 50,000-100,000 new users can be no heritable inside some of days. In fact, such ranking fraud raises nice issues to the mobile App business. As an example, Apple has warned of cracking down on App developers United Nations agency commit ranking fraud [2] within the Apple's App store. Within the literature, whereas there are some connected work, like internet ranking spam detection, on-line review spam detection, and mobile App recommendation, the matter of police work ranking fraud for mobile Apps remains under-explored. To fill this important void, during this paper, we tend to propose to develop a ranking fraud detection system for mobile Apps. On this line, we tend to determine many necessary challenges. First, ranking fraud doesn't continuously happen within the whole life cycle of associate degree App; therefore we want to notice the time once fraud happens. Second, because of the massive variety of mobile Apps, it's tough to manually label ranking fraud for every App, therefore it's necessary to possess the simplest way to mechanically notice ranking fraud while not victimisation any benchmark data. Finally, because of the dynamic nature of chart rankings, it's difficult to spot and make sure the evidences connected to ranking fraud. Indeed, our careful observation reveals that deceitful Apps don't continuously be hierarchal high within the leader board, however solely in some leading events, that kind completely different leading sessions. Note that we are going to introduce each leading events and leading sessions intimately later. In alternative words, ranking fraud typically happens in these leading sessions. Therefore, police work ranking fraud of mobile Apps is really to notice ranking fraud inside leading sessions of mobile Apps. Specifically, we tend to initial propose a straightforward nonetheless effective rule.

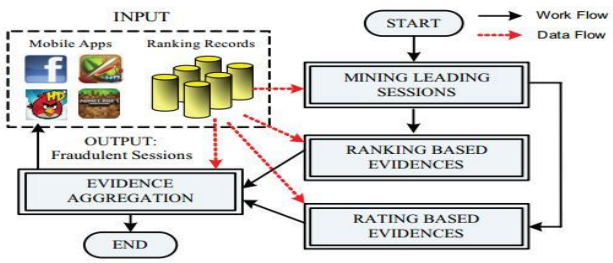


Figure 1: The framework of the ranking fraud detection system for mobile Apps.

To identify the leading sessions of every App supported its historical ranking records. Then, with the analysis of Apps' ranking behaviours, we discover that the dishonest Apps typically have completely different ranking patterns in every leading session compared with traditional Apps. Thus, we have a leaning to characterize some fraud evidences from Apps' ancient ranking records, and develop 3 functions to extract such ranking based mostly fraud evidences. Still, the ranking based mostly evidences is full of some legitimate selling campaigns, like "limited-time discount". As a result, it's not comfortable to solely use ranking based mostly evidences. Therefore, we have a tendency to more propose 2 functions to get rating based mostly evidences that mirror some anomaly patterns from Apps' historical rating records. Additionally, we have a leaning to develop secondary degree unsupervised proof aggregation system to integrate these 2 kinds of evidences for evaluating the quality of leading sessions from mobile Apps. Figure one shows the framework of our ranking fraud detection system for mobile Apps.

II. LITERATURE SURVEY

Leif Azzopardi et al. [2] studied AN work the connection between Language Model confusion and IR exactness Recall Measures the confusion of the language model incorporates a systematic relationship with the doable exactness recall performance though' it's not statistically important. A latent variable unigram based mostly luminous flux unit, that has been winning once applied to IR, is that the therefore known as probabilistic latent linguistics compartmentalization (PLSI). Ee-Peng Lim et al. given variety of detection Product Review Spammers exploitation Rating Behaviours to notice users generating spam reviews or review spammers. We have a tendency to establish much characteristic behaviour of review spammers and model these behaviours therefore on notice the spammers. David F. Gleich et al. [4] has done a survey on Rank Aggregation via Nuclear Norm reduction the method of rank aggregation is intimately tangled with the structure of skew-symmetric matrices. To produces a brand new technique for ranking a collection of things. The essence of our plan is that a rank aggregation describes a partly stuffed skew-symmetric matrix. We have a tendency to extend AN algorithmic rule for matrix completion to handle skew-symmetric information and use that to extract ranks for every item. Alexandre Klementiev, Dan writer et al. [9] studied AN unattended Learning algorithmic rule for Rank Aggregation; (ULARA) that returns a linear combination of the individual ranking functions supported the principle of rewardable ordering agreement between the rankers.

III. SYSTEM OVERVIEW

Detection of ranking fraud for mobile Apps continues to be below an issue to analysis. To fill this significant lack, we have a tendency to propose to develop a ranking fraud detection system for mobile Apps. We have a tendency to additionally verify many necessary challenges. 1st challenge, within the whole life cycle of Associate in Nursing App, the ranking fraud doesn't forever happen, therefore we'd like to discover the time once fraud happens. This challenge is thought-about as police investigation the native anomaly in situ of worldwide anomaly of mobile Apps. Second challenge, it's necessary to own a climbable thanks to completely discover ranking fraud while not exploitation any basis info, as there are a unit immense range of mobile Apps, it's terribly tough to manually label ranking fraud for every App. Finally, thanks to the dynamic nature of chart rankings, it's tough to search out and verify the evidences related to ranking fraud that motivates USA to find some implicit fraud patterns of mobile Apps as evidences.

MODULES:

Module 1: Leading events Given a positioning limit K nine two [1, K] a main occasion e of App a contains a amount vary conjointly, relating rankings of a, Note that positioning edge K * is applied that is generally littler than K here on the grounds that K is also immense (e.g., quite one,000), and also the positioning records past K _ (e.g., 300) don't seem to be exceptionally useful for recognizing the positioning controls. Moreover, its finding that a number of Apps have a number of closes driving even that square measure close to each other and structure a main session.

Module 2: Leading Sessions Instinctively, in the main the leading sessions of mobile app signify the amount of recognition, and then these leading sessions can comprise of ranking manipulation solely. Hence, the difficulty of characteristic ranking fraud is to spot deceptive leading sessions. In conjunction with the most tasks is to extract the leading sessions of a mobile App from its historical ranking records.

Module 3: characteristic the leading sessions for mobile apps primarily, mining leading sessions has 2 sorts of steps regarding with mobile fraud apps. Firstly, from the Apps historical ranking records, discovery of leading events is finished then second merging of adjacent leading events is finished that appeared for constructing leading sessions. Certainly, some specific formula is incontestable from the pseudo code of mining sessions of given mobile App which formula is in a position to spot the sure leading events and sessions by scanning historical records one by one.

Module 4: characteristic evidences for ranking fraud detection Ranking based mostly proof it concludes that leading session contains of assorted leading events. Thus by analysis of basic behaviour of leading events for locating fraud evidences and conjointly for the app historical ranking records, it's been determined that a selected ranking pattern is usually happy by app ranking behaviour during a leading event. Rating based proof previous ranking based evidences square measure helpful for detection purpose however it's not decent. Breakdown the matter of "restrict time reduction", identification of fraud evidences is planned thanks to app historical rating records. As we all know that rating is been done when downloading it by the user, and if the rating is high in leader board significantly that's attracted by most of the mobile app users. The ratings throughout the leading session provides rise to the anomaly pattern that happens throughout rating fraud. These historical records will be used for developing rating based mostly evidences.

Review based proof we tend to square measure acquainted with the review that contains some matter comments as reviews by app user and before downloading or mistreatment the app user mostly opt to refer the reviews given by most of the users. Therefore, though thanks to some previous works on review spam detection, there still issue on locating the native anomaly of reviews in leading sessions. Therefore supported apps review behaviours, fraud evidences square measure accustomed notice the ranking fraud in Mobile app.

IV. EVIDENCE AGGREGATION

A. Algorithms

Algorithm 1

Proposed Algorithm: K-Means
Input: ratings & rankings in Historical records
Output: removes fraud rating & ranking

- Take 2 clusters randomly
- Find the Mean
- Group the nearest points to the mean
- Find the Mean of the formed group
- Based on Mean find the nearest numbers and form the group
- Repeat the process until there is no exchange of data
- If the Mean comes same as consequently stop the process
- Choose the cluster which has more no.of ratings and remove the other cluster which has fraud ratings
- Calculate the mean of cluster and take it as evidence score.

Method 1

Weighted Ranking Method

- .Take evidence scores
- Take any Parameter and Based on that parameter give weights to the apps
- Multiply weights with evidence Scores
- We will get Final Evidence scores
- Based on the Final Evidence Score Rank is given to app.

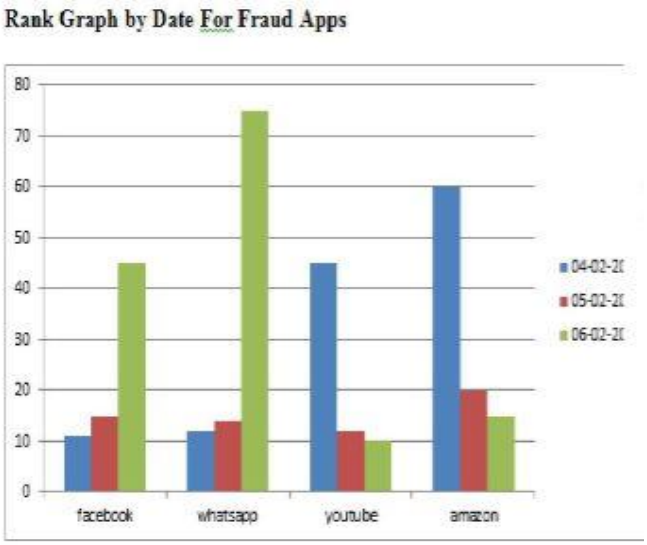
Advantages:

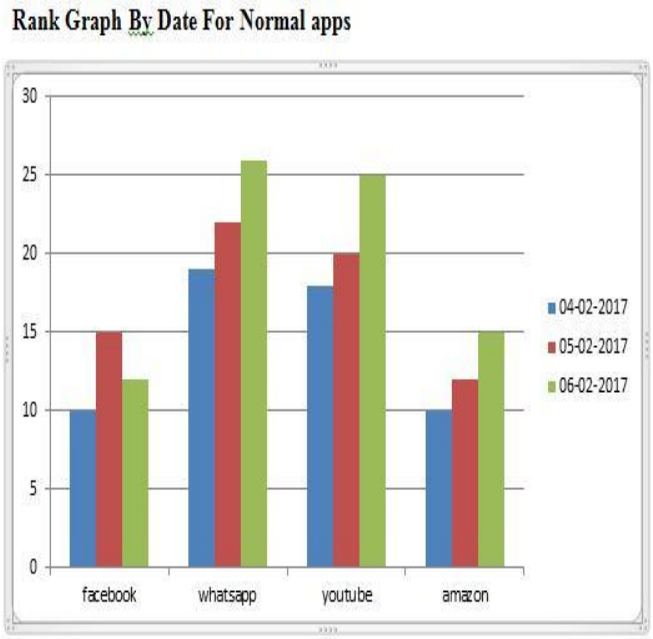
K-means clustering is very Fast, robust and easily understandable. If the data set is well separated from each other data set, then it gives best results. The Advantage of K-Means is to find the fraud ratings and reviews and remove from the historical records.

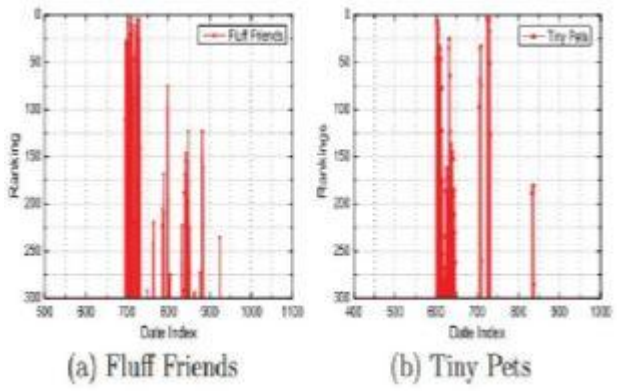
By using weighted ranking method we can get the rankings of apps.

Result Graphs:

Here, main attention is on extracting different evidences such as reviews, ratings, ranking and download information from historical records of data set. Data set contains the historical reviews, ratings of mobile apps. In the result parts calculates and merge the evidences with help of evidence aggregation method.







Each new test App can reuse this model for detecting ranking fraud. Specifically, the learnt weight parameters (i.e., wi) in our approach EA-RFD are 0.24 (0.22), 0.30 (0.28), 0.19 (0.18), 0.15 (0.17), and 0.12 (0.15) for each evidence in Top Free 300 (Top Paid 300) data set, respectively. It indicates that ranking based evidences are more effective than rating

based evidences. However, it is not clear how many learning data are required? To study this problem and validate the robustness of our approach, we first rank all leading sessions by modelling with weight parameters learnt from the entire data set. Then we also rank all leading sessions by modelling with weight parameters learnt from different segmentation of the entire data set (i.e., 10%,..., 100%). Finally, we test the root mean squared error (RMSE) of the ranking of leading sessions between different results. Figure 2 shows the results of robust test on two data sets. We can find that the aggregation model does not need a lot of learning data, thus the robustness of our approach is reasonable.

V. CONCLUSION

In this paper, we tend to develop a ranking fraud detection system for mobile Apps. Specifically, we tend to initial showed that ranking fraud happened in leading sessions and provided a way for mining leading sessions for every App from its historical ranking records. Then, we tend to known ranking based mostly evidences and rating based evidences for police investigation ranking fraud. Moreover, we tend to projected an improvement based mostly aggregation methodology to integrate all the evidences for evaluating the quality of leading sessions from mobile Apps. A distinctive perspective of this approach is that each one the evidences will be modelled by applied math hypothesis tests, so it's simple to be extended with different evidences from domain data to notice ranking fraud. Finally, we tend to validate the projected system with intensive experiments on real-world App information collected from the Apple's App store. Experimental results showed the effectiveness of the projected approach.

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] 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.

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

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

[7] L. Azzopardi, M. Girolami, and K. V. Risjbergen. Investigating the relationship between language model perplexity and ir precisionrecall measures. In Proceedings of the 26th International Conference on Research and Development in Information Retrieval (SIGIR'03), pages 369–370, 2003.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan. Lantent dirichlet allocation. Journal of Machine Learning Research, pages 993–1022, 2003.

[9] Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou. A taxi driving fraud detection system. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining, ICDM '11, pages 181–190, 2011.

[10] 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.

