International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Detection Of Fraud Ranking For Mobile App Using IP Address Recognition Technique

Esther Nowroji. S[1], Vanitha. S[2]
[1]PG Scholer, [2]Assistant Professor, Department of Computer Science and Engineering
Dr. N. G. P Institute of Technology, Coimbatore, India

Abstract— fraudulent activity in mobile app market means intruders or app developers use shady means to increase their app rating to bring their app in top 20 list to inflate their app sale. Knowledge engineering domain usually uses the methodologies to extract the useful knowledge from the given large data. Ongoing rapid growths of online data have created the need of KDD. Also ongoing rapid growth of online rating and review system to the app, make fraud app has been launched in the mobile market and let them be downloaded and used by many users. The fraud mobile app is not worth to use and wasting device memory. Sometimes such app is created with malicious software which is harmful to the device. To avoid this situation the fraud app should be find out. In existing work, fraud rankings are detected by applying the mining algorithm in app review. Local anomaly was detected instead of global anomaly. The analysis had been done reported. Human evaluators evaluated and produce the result. Time complexity is more to evaluate. To overcome this drawback, FRDS is proposed. To detect the fraud app, app's reviews should be checked. To check whether the app reviews are fraud or not, the Fraud Ranking Detection System is proposed. In the mobile market, each mobile has its own unique IP Address. Hence, each user has unique IP Address. When giving the reviews to app, user IP Address is extracted by using the IP Address recognition technique. So that, from one IP Address number of reviews cannot be provided to same app. In this way fraud review is prevented in proposed work. This approach decreases the evaluation time of the result, hence it is efficient than the existing approach.
Index Terms — Detecting Fraud Ranking, Ratings and Reviews, Aggregation method

I. INTRODUCTION

Past view years more number of Apps has increased with their different updated version. In mobile App market, App store launch the daily App Leaderboard. Leader board shows the chart ranking of the populated Apps. App Leader board is one of the important way of promoting the mobile App. App which posses higher rank in a Leader board will get the huge download and leads to the million dollars of revenue. So, developers has taken many way to promote their App to make their App place in a higher position in a Leaderboard to increase the App downloads. Instead of follow Traditional marketing solution, App developers fall into the deceptive activities to boost App ratings which show the drastic increase in ratings of the App in Leaderboard. App developer manipulates chart ranking on an App store. This eventual manipulation is implemented with the help of method called Human Water armies to increase the App download and ratings in a short time of periods. There is an observation which reveals that always mobile Apps are not ranked at the high position in the Leaderboard. Leading events form the many leading sessions. In those leading events mobile Apps may have high position in the Leaderboard. Fraud ranking may happend in those leading sessions. Efficient algorithm is proposed to identify the Leading events and leading session depending on their historical records.

In the literature survey, there are many related works are available like web ranking spam detections [4], online review spam detections [6], mobile Apps recommendation [8]. Figure 1 represents the fraud detection system. There is a problem of fraud ratings detection is not having a solution. This paper proposes the solution for detecting the fraud ratings. Instead of finding the global anomaly, challenge is to find the local anomaly to determine the fraud raking. Leading events and Leading sessions find out for getting result. Usually Leaderboard is updated daily. Leading event is time range which indicates how long App holds the rank in a Leaderboard. These different leading events form the leading session. Mining is extracting the useful information from the given collection of data. Classification is the unsupervised Approach which does not need any labeled set of data. Data aggregation is one of the important concepts in the unsupervised learning Approach.

In a Leaderboard, dynamic data will be aggregated periodically using this one the aggregation method. Aggregation methods such as sum, avg, median, minimum, maximum, variance, std deviation, set, arb. MAC Address is a unique for each machine which never be changed. App fraud ranking detection system makes use of the MAC Address of user's source node. App which is

positioned in the higher rank will be checked for MAC address verification. This method verifies the relatively same reviews are from the same MAC address in particular period of time.

II. RELATED WORK

In mobile Application market, ranking fraud means fraudulent activities that push up Apps in popularity list. Application developer uses the shady means to increase their sales. To locate the ranking fraud, process of mining active period will take place. Leading sessions are active period. This Leading session emphasize on local anomaly not on global anomaly. There are three evidences are investigated. They are rating based evidence, ranking based evidence, review based evidence. Investigation[1] is done by modeling the App's rating, ranking and review behavior with the help of statistical hypothesis test. Optimization based aggregation method is used to integrate these three evidences with real world app data from iOS App store. To built a graph for reviews and reviewer [9] the following equation is used to determine the relationship between the two quantities x and y. The given below is called power law helpful to plot the graph for product reviews and for various reviewers. The Power law is written as

$$y = axk \qquad (1)$$

where *a* and *k* are constants. If we take the log on both sides, we will obtain a straight line on a log-log plot. Figure 4 represents the graph. Rank aggregation used to combine [11], [14] the rankings and produce the joint ranking. Without supervision votes from the respective rankers with domain-specific expert will be aggregated. Votes of the each judge is produced by the sampling model.

Web page content spam detection is focus on these two steps. Even though Crawlers give importance to the well connected and more important page, pages are ranked in higher position by the search engine. Spam pages are numbered Approximately [4], [18], [19] according to the perception of the users. Crawler estimate the impact of web spam discard the spammed web page which has been reported. Prevalence of spam relative to number of words on page represents more than half of all the web pages contain less than 300 words in that only 12.7% of all the pages contain at 1000 words at least. Poisson Distribution of the bar graph is estimated with a mode of 2 words, a median of 281 words, and a mean of 429.2 words.

Unsupervised learning Approach for Rank Aggregation(ULARA) make the data fusion real-world task from ad hoc retrieval system. Group of shared task N participants(50) are provided with Q number of Queries(50) and return the 1000 documents[12] which is relevant to those queries. ULARA is used to combine these rankings of each group into aggregate rank function R. Performance will be quantified by the mean average precision metric.

Two types of ranking systems are used to overcome the hurdles in learning model. Depending upon the distance function definition overcome[2] the hurdles. Two types are permutation and top-k list. Further individual judges specify ranking over the k objects out of n. For example, top-10 list items could be associated with the 10 items in the first page of a result produced by the web search engine. Decomposability property will be satisfied for permutation as well as top-k list which estimate LHS efficiently.

In many Applications, rank aggregation is used like Genome Database Construction, Document Filtering, Database Middleware Construction[5], [21], [22] Spam Webpage detection, Word association finding, Multiple search and Similarity Search. The main aim of the rank aggregation is to assigning the Real-valued score to individual entities by aggregating the every ranking provided by the base ranker. Sort the entities upon their score without change in generality. Median Rank aggregation sorts the entities upon the median of the rank in ranking list.

III. PROPOSED WORK

Latent relationship between three evidences is low. Entities used are not correlated with each other. Hence latent relationship is lost. Variable which does not have any correlation between them cannot result in a latent construct upon the common factor model. Huge number of observable variable cannot be aggregated in this model to understand the data. System will not work for dynamic data. Existing system has collected the real-world data set from iOS App store. Estimation process has been done on these real-world data set.

A. Fraud Ranking Detection System

In fraud detection system, App will be rated and reviewed by the user. Proposed system uses the rank aggregation method which aggregates the ratings given for particular App given by the user. Aggregate method usually does the aggregation with its predefined methods. If the aggregated ratings go beyond threshold value, system will check for MAC of corresponding user who gave the

rating. If more than one rating from same MAC within a certain time period, that rating will be called as fraud rating. Those ratings will be deleted from aggregated ratings in App leader board. This is proposed for Android Operating System which is popularly used by maximum number of users worldwide. System focus on issue of shady means which means that App is promoted by fake ratings. In mobile Application market this is big issue. Very few research works has been done for this issue. When an App was promoted by ranking manipulation it will be the top in the market, other users will buy the App. This activity affects other App reputation and some legitimate marketing campaigns, such as limited time discount.

With this modern world, many smart phone users are download Apps from internet. Money motive App developers release their App and push them to top 25 lists to increase the view and downloading to yield more money. App with ranking manipulation always has an expected ranking target and the hired marketing firms also charge money can be earned. After reaching, maintaining the expected ranking for a required period, the manipulation [19] will be stopped and ranking of the malicious App will be decreased dramatically. As a result, the suspicious leading event may contain very short rising phase, recession phases.

Meanwhile, cost of ranking manipulation with high ranking expectations is quite expensive due to the unclear ranking manipulation of App store and the fierce competition between App developers. Therefore, the leading event of fraudulent apps have very short maintaining phase along with high ranking positions. Generating graph which helps in finding the fraudulent ratings for set of 10 Apps. According to the rating given daily, ranking is done and show it in the graphical representation. In graph there are two variant bars. One represents the how many user rate the App in hours. Second represent the how many users rate the App in slow dating (consecutive date). This Emphasis on finding the fraud rating when the first bar is greater than second bar. The proposed framework is scalable and can be extended with the other domain generated evidences for ranking fraud detection. Finally, system evaluate the proposed system with data collected as dynamic ratings for App in daily basis. Efficiently utilize the aggregation method in the Fraud Detection System. The main aim of this aggregation method is to collect real-world data and make the data fusion. Easy accessibility of the Fraud Detection System is one of important advantage. The system provides the well security for the user to access the user interface.

B. Aggregation

For fraud ranking detection, rank of each App is aggregated individually. This aggregation method is having many ways. Some methods are focusing on learning global ranking for every candidate. But this is not a proper method to detect the fraud ranking for new App. Another method is supervised learning Approach which needs the labeled training data. This system proposes the unsupervised learning Approach. System defines the final ranking aggregation score Ψ_(h) as linear combination of every the existing ranking as Equation 2.

$$\Psi\_(h) = \sum_{i=1}^{n} w_i \times \Psi_i(h), \; s.t. \sum_{i=1}^{n} w_i = 1 \qquad (2)$$

Where n=8 is the number of ranking for App, weight wi ∈ [0, 1] is the aggregation parameter of the ranking Ψi(h). First system proposes the assumption called Principle 1 for the ranking aggregation Approaches. System assumes that effective rankings must have similar ranking scores for each leading session, when poor ranking will generates different scores from others. Ranking will be consistent with plurality of the rankings would be given higher weight and ranking which tends to disagree would be given smaller weight.

for each ranking score $\Psi_i(h)$, system will measure the consistence using variance-like measure

$$\sigma_i(h) = (\Psi_i(h) - \Psi(h))2 \qquad (3)$$

where $\Psi(s)$ is the average ranking score of leading session s got from all N rankings.

$$\arg\min_{\mathbf{w}} \sum_{a2A} \sum_{s2a} \sum_{i=1}^{n} w_i \cdot \sigma_i(h) \qquad (4)$$

$$s.t \sum_{i=1}^{n} w_i = 1; \; \forall w_i \geq 0 \qquad (5)$$

This system exploits gradient based Approach with exponentiated updating to resolve the problem. System assigns *wi* = 1 to *n* as initial value. After that for each h, system will compute the gradient.

C. Text Mining Algorithm

Normal app always contains had not unique reviews because reviewers have the different personal opinion and usage experiences. To address the fraud review, mutual similarity between the reviews within the particular leading session should be found. Reviews contain sentence from which stop word will be removed and those reviews will be normalized.(e.g., “good”, ”best”).

1: $E_s = \emptyset$; $e = \emptyset$; $s = \emptyset$; $t^e_{start} = 0$;
2: **for each** $i \in [1, |R_a|]$ **do**
3: **if** $r^a_i \leq K^*$ **and** $t^e_{start} == 0$ **then**
4: $t^e_{start} = t_i$;
5: **else if** $r^a_i > K^*$ **and** $t^e_{start} \neq 0$ **then**
6: //found one event;
7: $t^e_{end} = t_{i-1}$; $e =< t^e_{start}, t^e_{end} >$;
8: **if** $E_s == \emptyset$ **then**
9: $E_s \cup = e$; $t^s_{start} = t^e_{start}$; $t^s_{end} = t^e_{end}$;
10: **else if** $(t^e_{start} - t^s_{end}) < \phi$ **then**
11: $E_s \cup = e$; $t^s_{end} = t^e_{end}$;
12: **else then**
13: //found one session;
14: $s =< t^s_{start}, t^s_{end}, E_s >$;
15: $S_a \cup = s$; $s = \emptyset$ is a new session;
16: $E_s = \{e\}$; $t^s_{start} = t^e_{start}$; $t^s_{end} = t^e_{end}$;
17: $t^e_{start} = 0$; $e = \emptyset$ is a new leading event;
18: **return** S_a

Fig.1. Text mining Algorithm

The fig.1 is representing the Text mining algorithm which is used to mine the local anomaly from reviews of users. Those normalized words called as the local anomaly which will be further used for human evaluation of final results which is analyzed and reported in the existing work of the fraud ranking detection system.

IV. SYSTEM MODEL

A. IP Address Recognision Technique

In mobile app market, each mobile device has its own IP Address. Single user environment contain unique IP Address connected to a network which does not having duplicate within their intra network area. Once user entering the ratings to the app through online, IP Address is extracted from user network and those are stored in Database for further mapping. If the user’s IP Address is already there in the Database those ratings from that IP Address will not be included with the app rating.

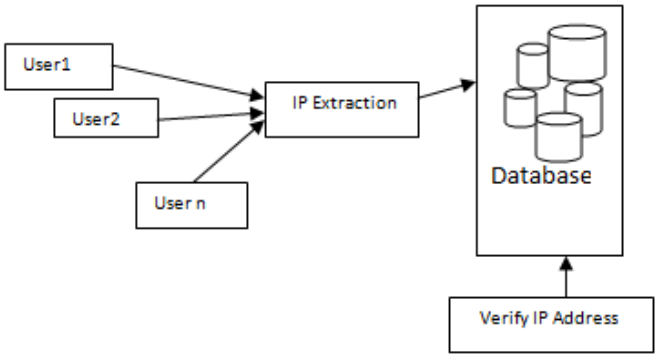

Fig.2. IP Address Extraction Process

B. Rank Aggregation With Mined Data

After IP Address extraction, filtered ratings are further ranked according to the aggregation methods. Aggregation methods contain

many default functions. Among those function avg() function used to aggregate the ratings to rank the app in the leaderboard. Finally as analysis work had done on the aggregated ratings and ranking, the evaluation is done along with the normalized review called local anomaly.

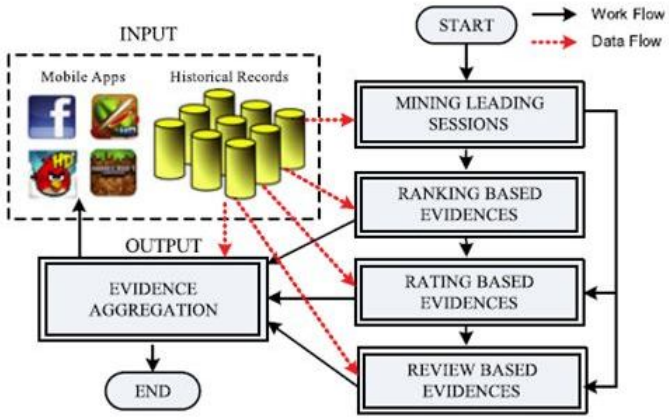

Fig.3. Fraud Ranking Detection System

V. EXPERIMENTAL RESULT

A. *Experimental Data*

Experimental data are collected from the dynamic ratings provided by the reviewers. Ratings are stored in the MySQL database which is one of the flexible databases. Entered ratings are aggregated using the methods implemented in the system. While providing the ratings to each app using the IP Address recognition technique user IP Address is extracted and stored in the database as table. Each time user enter the ratings to app, their IP Address is map with the IP Address table in database. Whether the user IP Address is already exist in the table, that entry will not be added to overall ratings to do the aggregation operation.

B. *Experimental Setup*

In existing system dataset is collected from the ios app store of two years. On that data set analysis work was done. Result was evaluated by the human evaluators. For each user evaluation process was done by the human evaluators. It took lots of time. To overcome this problem only IP Address extraction of single user is proposed. It is worked out as defined above. After IP Address verification, true ratings are filtered and Ranked according to aggregation methods through the avg() function.

VI. CONCLUSION & FUTURE ENHANCEMENT

In mobile app market, the term called fraud app is getting popular. In these days detection and prevention are playing vital role in mobile market. For the detection of fraud review to the single user system (i.e., mobile), the Fraud Ranking System is proposed. Ratings are aggregated to provide rank for each app. To whether the reviews are fraud or not, system used the IP Address recognition technique. The reviewer's source address is verified for the uniqueness. The user will be blocked when their IP Address is existing in the IP Address table, that review is removed from the overall ratings of the app. Proposed system prevented from provision of reviewers fraud ratings to a mobile app. Fraud Ranking Detection System does not need any human evaluation. So finally time complexity is reduced with Fraud Ranking Detection System. System has worked efficiently with dynamic data. Fraud Rank Detection System is used for single user system like mobile which has unique IP Address. Due to the advancement in network technology, there were many security attack techniques under usage nowadays. One of the security attack technique is IP snooping which means that user can change their IP Address. By this, User can change their IP Address and give the ratings to same app many times. To get the true value Detection system should be enhanced for further techniques which will produce the true value.

REFERENCES

[1] H. Zhu, H. Xiong, Y. Ge, and E. Chen. Ranking fraud detection for mobile Apps: A holistic view. In Proceedings of the 23rd ACM international conference on Information and knowledge management, CIKM '14, Vol 27, No. 1, Jan 2015.

[2] H. Zhu, H. Xiong, Y. Ge, and E. Chen. Ranking fraud detection for mobile Apps: A holistic view. In Proceedings of the 22nd ACM international conference on

Information and knowledge management, CIKM '13, 2013.

[3] 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.

[4] 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.

[5] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly. Detecting spam web pages through content analysis. In Proceedings of the 15th international conference on World Wide Web, WWW '06, pages 83–92, 2006.

[6] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li. Supervised rank aggregation. In Proceedings of the 16th international conference on World Wide Web, WWW '07, pages 481–490, 2007.

[7] 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.

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

[9] 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.

[10] N. Jindal and B. Liu. Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM '08, pages 219–230, 2008.

[11] G. Heinrich. Paramter stimaion for text analysis. Technical report, University of Lipzig, 2008.

[12] A. Klementiev, D. Roth, K. Small, and I. Titov. Unsupervised rank aggregation with domain-specific expertise. In Proceedings of the 21st international jont conference on Artifical intelligence, IJCAI'09, pages 1101–1106, 2009.

[13] 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.

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

[15] 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.

[16] 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.

[17] T. L. Griffiths and M. Steyvers. Finding scientific topics. In Proc. of National Academy of Science of the USA, pages 5228–5235, 2004.

[18] G. Shafer. A mathematical theory of evidence. 1976.

[19] K. Shi and K. Ali. Getjar mobile application recommendations with very sparse datasets. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '12, pages 204–212, 2012.

[20] M. N. Volkovs and R. S. Zemel. A flexible generative model for preference aggregation. In Proceedings of the 21st international conference on World Wide Web, WWW '12, pages 479–488, 2012.

[21] 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

[22] S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '12, pages 823–831, 2012.

[23] B. Yan and G. Chen. Appjoy: personalized mobile application discovery. In Proceedings of the 9th international conference on Mobile systems, applications, and services, MobiSys '11, pages 113–126, 2011.

[24] G. Heinrich. Paramter stimaion for text analysis. Technical report, University of Lipzig, 2008.