Pak. J. Biotechnol. Vol. 13 (special issue on Innovations in information Embedded and Communication Systems) Pp. 5-9 (2016)

IDENTIFYING RATING AND REVIEW BASED RANKING FRAUD IN MOBILE APP MARKET

Uthra, K[1] and R. Dhanalakshmi[2]

Dept. of computer science and Engineering, RMK Engineering College, *Kaverapettai,* Gummidipoondi taluk, thiruvallur,India Email[1]: Uthrak998@gmail.com, Email[2]: rdl.it@rmkec.ac.in

ABSTRACT

Nowadays ranking fraud in mobile App market became more popular in the market in order to display their apps in the popularity list and to boost their sales.As the applications increase the fraud is also increasing. Therefore, there will be an increase in ranking fraud in the upcoming time, as the number of Apps developers and applications will likely grow very significantly. Many traditional methods of fraud analysis have been used to detect fraud. But these methods are complex and time-consuming. However, there is more need to adopt some better techniques which can ensure the ranking fraud detection efficiently by data mining analysis. This paper explores the data mining methods to identify the fraud by using Rank Aggregation algorithm. Further three types of proofs are studied they are ranking, rating and user comment proofs. And an aggregation method is used to aggregate all the proofs and will produce an optimized report for fraud detection.

Keywords: Mobile Apps market, fraud detection, aggregation method, records of the apps, rating and user comments proofs.

INTRODUCTION

For the past few years, the number of mobile Apps increases day by day. As per the statistics taken in 2012, per day 1 millions of apps been downloaded in which 91% are free apps and the remaining 9% are paid one. And in order to increase the sales, each app is ranked and maintained in the leading board. Many fraudulent activities taking place on this board only. Thus, leading sessions are used for marketing the apps. Higher the rank usually lead to huge downloads of the app and billion of rupees in gain. And they also follow various ways to move their app to a higher position and will also undergo various marketing techniques including advertisements. Recently many trends are being used in order to boost the apps sales.

For example, Google found the developer who undergone this type of frauds in their apps store.

Some related works, such as Latent Dirichlet allocation, A taxi driving fraud detection systemic in city taxis, Rank aggregation via nuclear norm minimization, An unsupervised learning algorithm for rank aggregation, Unsupervised rank aggregation with distance-based models the problem of detecting ranking fraud for mobile Apps is still under-explored are used for literature works. Thus, we proposed a paper for ranking mobile fraud detection.

To solve this problem, first local deviations are proposed later they are combined and proposed for the global deviations. Later, all these proofs are categorized and it is maintained in the database. And also, the patterns which are followed by the fraudulent apps to rank their apps are unique that is, it varies for each and every leading session of the apps stores. The ranking is calculated based on the user comments and rating. Thus, further two types of proofs are recorded which helps to detect the fraud in the leading board. Finally, an unsupervised proof is produced and an aggregation method is used to combine proofs for ranking the believability of the session from Apps. Thus, our proposed system is more scalable, efficient and its performance is high comparing the existing.

Future, this paper is divided into following. Part 2 report about the leader board and rank aggregation. Part 3 holds the planned approach and results are discussed in part 4. Finally, part 5 describes the conclusions and future works.

1. RELATED WORKS:

In this part, few related works of the leader board session and rank aggregation are discussed.

2.1 Leading board session

It is observed that finding the historical records of the Apps and which helps to reduce the ranking fraud. And moreover now all the Apps are not placed higher in the leadersession but only in few cases. Thus, it is also found that there exist some beside sessions where they are closer to one another to form board sessions. Further, to find the ranking fraud from several leading sessions, an effective approach is developed called as unsupervised Evidences Aggregation based Ranking Fraud Detection (UEA-RFD).Specifically, this method is denoted by result based aggregations (i.e., first idea) as UEA-RFD-1.0, and deal with unsupervised rank aggregation (i.e., second idea) as UEARED- 2.0.

1 (Leading board Event): The ranking threshold T 2.5; K, board events E of App contain a time period ranges Re ¼ ½te start and ends the ranking of A, which full fills the start and K< start _T, and end _ T_ < end þ1. Further, 8 to 2 starts; te end p1, we have raked_ K.

2 (Leading sessions): The session contain the period range Ts ¼ ½ and Ts starts; Ts end and a corresponding session fe1; eng, which satisfies ts start ¼ te1 start, Ts end ¼ ten ends and there are no other leading sessions that make Ts. While, 8i 3 ½1; np, we have teip1 start to tie and < T, T is the defined thresholds of events.

Figure 1 Distribution of ios and android apps in leading sessions

Fig 2 is a leader board of the apple apps store where it daily updates the ranking of each and every app in their store. And it also gives the details of the app which lose its rank and also the app which moves up in the rank list. The fraud is happening mainly in these types of the leader boards and the proposed system helps to detect this fraud. Finally, an optimized report is generated which helps to detect the fraudulent apps.

Figure 2 Example of leader board

2.2 Rank aggregation

To define the ranking aggregation method, first need to define the goal. Here, the functions contain "super"- a list which would be as "close" as possible to all ordered list. The simple and effective requirements and our goal which made this more brief simple and efficient.

$$\Phi(\delta) = \sum_{i=1}^{m} w_i d(\delta, L_i),$$

This is in ordered form and of length k=jLij and W is the associated weight with list L, D corresponds to distance.

2. PROPOSED METHOD

Our research found that all the fraudulent Apps will have various ranking designs in their events as tested with other Apps. And here the client can directly communicate with the server and can get the information about the rank of the app. And the user is also having rights to rate and give comments about the apps in the leader board. And this rating and user comments are aggregated to produce the rank of the app. Here the exact time when the rating and comments were given is noted and maintained in the database. The request and response of the client and the server is secured using the private key.

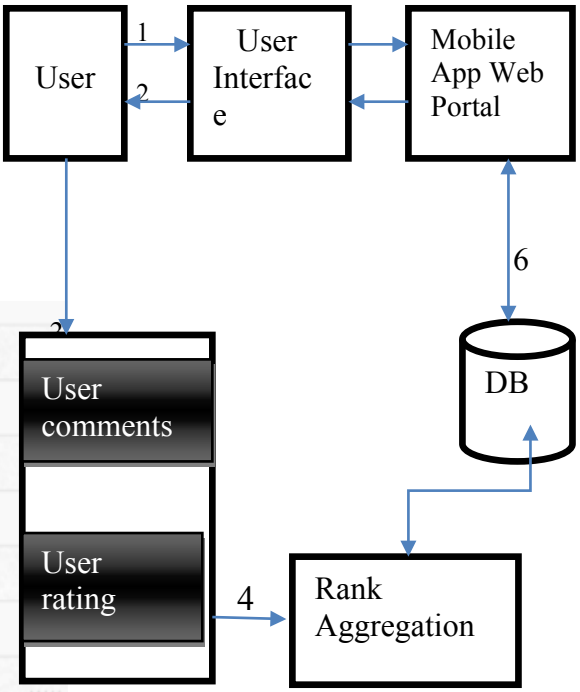

Figure 3 Architecture Diagram of Overall System

The figure shows the architecture of the overall system which consists of modules like User Interface, Identifying leading sessions, Ranking proof, Rating proof, User comments proof and Evidence aggregation. We Design new Client side page for actively interact with the server, for

improve secure purpose we generate unique Authentication scheme for each and every user(Application Data Owner)for check Competitor App Data Ranking History's. And also, we design another Client side page's for user search, user comments, rating, the App.

Table 1.Overview of the total number of apps

Number of apps in Google play store	1,600,030
Number of apps in windows phone store	341,000
Number of apps in apple store	100bn

3.1 User interface:

In This We Design new Client side page for actively interact with the server ,for improving secure purpose we generate unique Authentication scheme for each and every user(Application Data Owner)for check Competitor App Data Ranking History's.

And also, we design another Client side pages for user search, user comments, rating, the App.

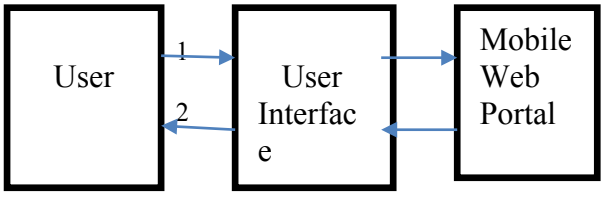

Figure 4 User interface diagram

3.2 *Identifying leading sessions*:

Raking fraud always happens in lead sessions. Thus, it means finding frauds in mobile Apps is detecting these frauds in the sessions of the App. Especially, here an effective fraud detecting algorithm is used to detect the fraudulent applications in the sessions. Then, with future research of the sessions, we found that all the fraudulent Apps will have various ranking designs in their events as tested with other Apps.

3.3 *Ranking Proof:*

The session is the combination of several small events and thus, first we analysed the most important underlying characteristic of the sessions in order to bring the proofs. By analysing the Apps records.

Table2. Statistics of the data an overview

Number of Mobile apps downloaded worldwide	102,082m
Projected number of apps downloads	258,692m
Number of free mobile apps	92.89bn
Number of paid mobile apps	9.29bn

We found that App's ranking in sessions will fulfil the basic design needs, where it holds varies stages they are, growing stage, protection stage and decline stage. Specially, in all sessions, each App's ranking will grows and reaches the peak for each time period T (growing stage) then comes the protection stage where it maintain its position and protect from other defects and finally comes the decline stage where the apps ranking will downturn and fall.

3.4 Rating proof:

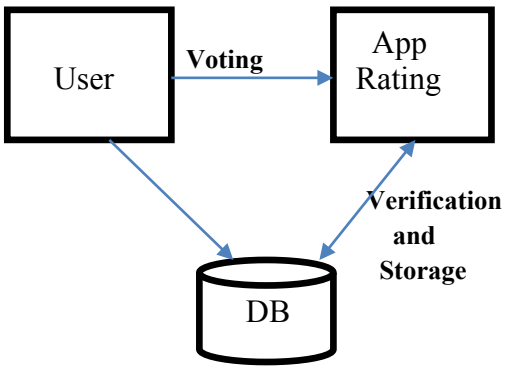

Figure 5 Rating based proofs diagram

The Ranking based proofs helps to detect the fraud in mobile apps but often this itself won't be sufficient in order to find the rating fraud the users comments will also play a vital role in this types of detection. Especially, after the launch of an application in their stores, the interested users will download these apps and share their experiences either by rating or by review and even by both review and rating. For App advertisement, the key thing is app rating. Higher the rating higher downloading and automatically will leads to more profit. It proofs that application rating is one of the important key in the fraud method. Certainly, if any apps undergone the fraud in the session can be found by its design and continues monitoring and by its previous records. Here all the ratings given by the users are recorded which includes the time, date and number of the mobile and these are verified with the historical records and then stored in the database.

3.4 User comments proof:

Apart from rating, almost all play stores allow the users to give reviews about the apps. Thus user comments help to find the user's personal view about the app. It proofs, the user comments is another vital key in ranking fraud. Specially, while downloading or purchasing a new mobile App, user will go through the apps historical user comments which help to take a better decision more positive comments will increase the sales of the app. Thus, imposter's often post fake comments in the sessions in order to inflate Apps sales. And this also helps to

increase the apps rank high in the sessions. However the older fraud detection method is not much effective. Here by using the effective algorithm and all the comments are recorded and these are verified using the stored comments and later these comments are stored in the database.

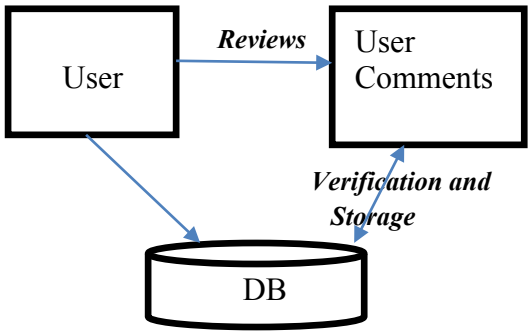

Figure 6 User comments based proofs diagram

3.5 Aggregating the proofs:

As final, we need to combine these proofs in order to find the ranking fraud in mobile apps. However at present lots of aggregation methods are used they are permutation based models, score based models and Dempster-Shafer rules. However, many methods are there but proper method for fraud detection is not found. Thus in order to solve this problem our proposed function is more effective. And also, we propose an unsupervised approach which also combines these proofs.

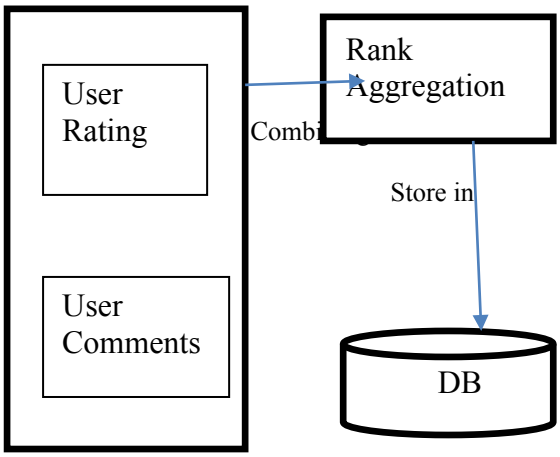

Figure 7 Aggregating the proofs diagram

4. DISCUSSION

Our key idea is to extract the proofs of the previously stored records. And this records holds the details about the user rating and the comments which helps to detect the fraud.

The proposed system helps both for local and global anomalies. Specifically, an App fraud scores for each Apps are given and the fraudulent apps is identified.

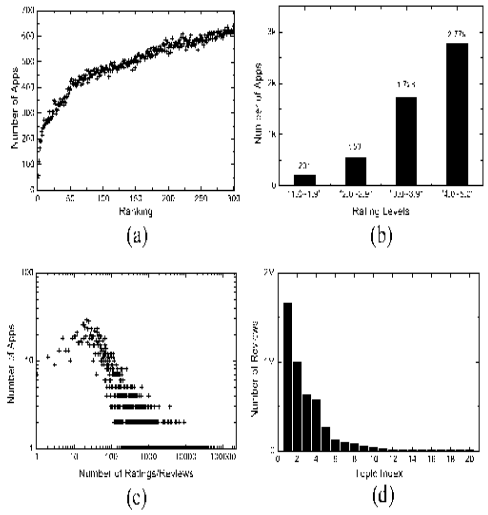

Figure 8 Data distribution in top 300 paid apps

However, our approach is scalable for integrating other proofs, such as the proofs based on the downloading information and App developers' reputation. Second, the proposed approach can detect ranking fraud happened in Apps' historical leading sessions also.

5. CONCLUSION AND FUTURE WORKS

All the issues regarding the current version of apps can be undertaken by using Mining Leading Session algorithm. And future the replicated version is identified by the admin by using Historical Records. The admin will also see the date of publication of the apps and the time, date exactly when the app is rated and commented. When the apps are detected as fraudulently published by the admin then the respective app will be blocked and a detailed report will be produced. The user can give the feedback at only once. Sentiword dictionary is used for finding the exact user comments. The admin can block the fake application. The User comments or Rating or Ranking given by users will be Correctly Calculated. Hence, a new user who wants to download an app for their own purpose can get a clear detail about the remaining applications. In the future, we plan to study more effective fraud proofs and analyse the latent relationship among rating, user comments and rankings for mobile apps.

6. REFERENCES

[1] 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).

[2] D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclear norm minimization," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining Pp. 60–68. (2011),

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

[4] Klementiev, D. Roth, and K. Small, "An unsupervised learning algorithm for rank aggregation," in Proc. 18th Eur. Conf. Mach. Learn. Pp. 616–623 (2007).

[5] Klementiev, D. Roth, and K. Small, "Unsupervised rank

aggregation with distance-based models," in Proc. 25th Int. Conf. Mach. Learn. Pp. 472–479 (2008)..

[6] Klementiev, D. Roth, K. Small and I. Titov, "Unsupervised rank aggregation with domain-specific expertise," in Proc. 21st Int. Joint Conf. Artif. Intell. Pp. 1101–1106 (2009).

[7] 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. Pp. 939–948 (2010).

[8] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, and H. Li, "Supervised rank aggregation," in Proc. 16th Int. Conf. World Wide Web Pp. 481–490 (2007),

[9] Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints,"in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Pp. 632–640 (2013).

[10] Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in Proc. 15th Int. Conf.World Wide Web Pp. 83–92 (2006).

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

[12] K. Shi and K. Ali, "Getjar mobile application recommendations with very sparse datasets," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Pp. 204–212 (2012).

[13] N. Spirin and J. Han, "Survey on web spam detection: Principles and algorithms," SIGKDD Explor. Newslett. 13 (2) 50–64 (2012).

[14] M. N. Volkovs and R. S. Zemel, "A flexible generative model for preference aggregation," in Proc. 21st Int. Conf. World Wide Web Pp. 479–488 (2012)