International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 01
January 2017

# Fraud Ranking Detection for Mobile Apps

[1] M.Archana,[2] T.Thirupathi
[1]PG Scholar, S.E., SR Engineering college,Warangal,Telangana, India.
[2]Assistant professor, Dept.of CSE,SR Engineering college, Warangal, Telangana, India.

**Abstract**: The number of mobile Apps has grown at enormous price during the last few decades. Ranking fraud in the mobile App market refers to fraudulent or fake pursuits which have a rationale of striking up the Apps within the fame record. It makes usual for App developers to post fake App ratings, to commit ranking fraud. Even as the value of stopping ranking fraud has been generally recognized, there's constrained understanding and study in this discipline. To this end, in this paper, we provide a quick view of ranking fraud and advise a ranking fraud detection process for mobile Apps. Mainly, we first endorse to safely locate the ranking fraud through mining the lively durations through utilizing mining leading session algorithm. Additionally, we investigate three varieties of evidences, i.e., ranking based evidences, score based evidences and overview based evidences, by way of finding out ancient records. We used an surest aggregation system to integrate all the evidences for fraud detection.

**Key Words**: Mobile Apps, Ranking Fraud Detection, Evidence Aggregation, Historical Ranking Records.

## I. INTRODUCTION

The number of mobile Apps has grown rapidly during the last few years. For illustration, as of the end of 2014, there are greater than 13 million Apps at Google Play. To stimulate the development of mobile Apps, many App outlets launched daily App leaderboards, which exhibit the chart rankings of most popular Apps. Certainly, the App leader board is one of the major ways for promoting mobile Apps. A higher rank on the leaderboard in general results in a colossal number of downloads and million bucks in revenue. Accordingly, App developers tend to explore more than a few methods similar to commercial to advertise their Apps so as to have their Apps ranked as excessive as feasible in such App leaderboards. Nevertheless, as a up to date development, instead of relying on typical advertising solutions, some App developers motel to a few fraudulent manner to intentionally boost their Apps and manipulate the chart rankings on an App retailer. That is normally applied with the aid of utilising so-known as "bot farms" or "human water armies" to inflate the App downloads, rankings and reports in an extraordinarily short time. For instance, an editorial from VentureBeat reported that, when an App was promoted with the help of ranking manipulation, it would be propelled from quantity 1,800 to the highest 25 in Apple's high free leaderboard and extra than 50,000-a hundred,000 new users would be received inside a couple of days. In fact, such ranking fraud raises high-quality concerns to the cell App industry.

Along this line, we determine a few essential challenges. First, rating fraud does no longer consistently happen in the whole existence cycle of an App, so we have got to detect the time when fraud happens. Such challenge will also be regarded as detecting the regional anomaly as an alternative of worldwide anomaly of cell Apps. second, as a result of the colossal number of mobile Apps, it is complicated to manually label rating fraud for each App, so it is main to have a scalable way to mechanically become aware of ranking fraud without making use of any benchmark data. Sooner or later, because of the dynamic nature of chart rankings, it's not handy to identify and affirm the evidences linked to ranking fraud, which motivates us to notice some implicit fraud patterns of mobileApps as evidences. In this paper, we furnish a brief view of rating fraud and propose a ranking fraud detection system for mobile Apps. Certainly, we first recommend to effectively find the ranking fraud through mining the energetic intervals by means of making use of mining main session algorithm. Such main sessions can be useful for detecting the regional anomaly as a substitute of world anomaly of App rankings. Furthermore, we examine three varieties of

evidences, i.e., ranking founded evidences, ranking established evidences and assessment based evidences, with the aid of modeling Apps' ranking, score and evaluation behaviors through analyzing its ancient documents. We advocate an optimization established aggregation method to combine the entire evidences for fraud detection.

## II. RELATED WORKS

The first is about web ranking spam detection. Exceptionally, the web ranking unsolicited mail refers to any deliberate actions which carry to chose webpages an unjustifiable Favorable relevance or importance [3]. For example,Ntoulaset al. [3] have studied various features of content material-centered unsolicited mail on the web and presented a quantity of heuristic approaches for detecting content based junk mail. Zhou et al. [3] have studied the challenge of unsupervised web ranking junk mail detection. Particularly, they proposed an effective on-line hyperlink unsolicited mail and time period spam detection ways utilising spamicity.

Lately, Spirin and Han [5] have reported a survey on web junk mail detection, which comprehensively introduces the standards and algorithms in the literature. Surely, the work of web ranking unsolicited mail detection is usually founded on the evaluation of ranking concepts of search engines like google, like PageRank and question term frequency. That is distinctive from rating fraud detection for mobile Apps.

The second class is focused on detecting online evaluation junk mail. For illustration, Lim et al. [9] have identified a number of indicative behaviors of evaluate spammers and model these behaviors to detect the spammers. Wu et al. [7] have studied the challenge of detecting hybrid shilling assaults on score information. The proposed method is centered on the semi supervised finding out and can be used for safe product suggestion. Xie et al. [8] have studied the difficulty of singleton evaluate unsolicited mail detection. Exceptionally, they solved this quandary by way of detecting the co-anomaly patterns in more than one assessment based time sequence. Although some of above approaches can be utilized for anomaly detection from old score and review records, they aren't in a position to extract fraud evidences for a given time period (i.e., leading session).

Ultimately, the third class includes the reviews on cellular App advice. For example, Yan and Chen [11] developed a mobile App recommender procedure, named Appjoy, which is situated on user's App utilization records to build a alternative matrix rather of utilizing specific consumer scores. Also, to solve the sparsity difficulty of App utilization records, Shi and Ali [4] studied a couple of advice items and proposed a content material based collaborative filtering model, named Eigenapp, for recommending Apps of their website Getjar. In addition, some researchers studied the quandary of exploiting enriched contextual information for mobile App recommendation. For illustration, Zhu et al. [10] proposed a uniform framework for customized context-mindful advice, which can integrate both context independency and dependency assumptions. However, to the exceptional of our data, none of previous works has studied the hindrance of ranking fraud detection for mobile Apps.

## III. PROPOSED METHOD

The mobile industry is growing rapidly, subsequently the number of mobile apps coming in the market is also increasing. As there are many apps available in market users are confused while downloading the apps for their use. They check the daily app leader boards for selecting app. But few fraudulent app developers are using shady means for bumping up their apps on the leader board in order to get revenue. So detect such fraud apps we develop a system based on evidences i.e. Ranking fraud detection using opinion mining for mobile apps.

As there is increase in the number of mobile apps, fraudulent Apps must be detected; we have proposed a simple and effective algorithm for identifying the leading sessions of each App based on its historical ranking of records. With the analysis of ranking behaviors of Apps, we recognize that the fraudulent Apps often having different ranking patterns in their each leading session compared with normal Apps. Some fraud evidences are identified from Apps" historical ranking records resulting in development of three

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 01
January 2017

functions to detect likewise ranking based fraud evidences.

Moreover, two types of fraud evidences based on Apps" rating and review history are proposed in Fig.1 depicts the framework of ranking fraud detection system for mobile Apps.

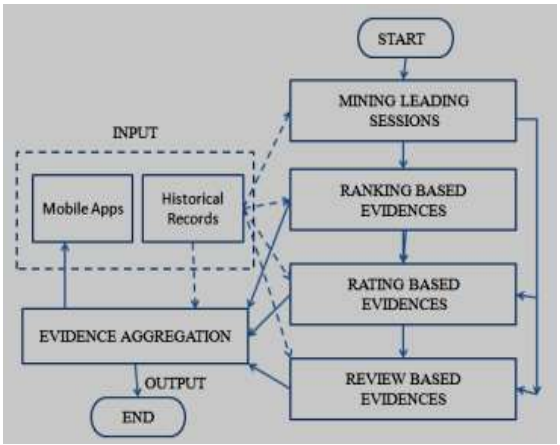


Fig. 1. Ranking Fraud Detection using Opinion Mining for Mobile Apps Overview

### A. Rating based evidences

Rating to app is given by the user who downloaded it, specifically after the app is published in the market. Hence rating is one of the main evidence in ranking fraud of apps.In this module it performs preprocessing of ratings that is it removes ratings that are less than or equal to two and calculates rating score by by summing all the ratings class collected and decision is taken on the basis of rating which scores high amongst all.

### B. Review based evidences

Reviews are familiar to all which provides the way for app user to write some textual comments regarding the personal experience of usage of that particular app.Therefore, manipulation of reviews is one way used by shady app developers to promote their app. Hence reviews are used to detect the ranking fraud in Mobile App industry. This module performs pre-processing of reviews and then performs sentiment analysis on pre-processed reviews.It will find out whether the comment is positive, negative or neutral. If word is positive then it will add plus one to score if word is negative it will minus one from score. Sometimes it is unable to find sentiment of some reviews, that time it makes the use of Naïve Bayes classifier.In this way it will find final score by analyzing sentiment of each review and determine whether app is fraud or not on the basis of review evidences.

### C. Ranking based evidences

As per the observation the mobile apps does not always ranked high in the leaderboards, in fact in some leading events only. Further, App having adjacent leading events are merged to form leading sessions. Hence, the problem of identifying ranking fraud is to find out vulnerable leading sessions. There are two phases for mining leading sessions. Firstly, we need to discover the leading events from the historical ranking records of apps. Secondly, merging of adjacent leading events must be done for constructing leading sessions.

Specially, Algorithm 1 demonstrates the pseudo code of finding leading sessions for a given App „a" is.

Algorithm 1 Mining Leading Sessions
Input 1: a's historical ranking records $R_a$;
Input 2: the ranking threshold K*;
Input 3: the merging threshold $\phi$;
Output: the set of a's leading sessions $S_a$;
Initialization: $S_a = \emptyset$

1: $E_a = \emptyset$; e = $\emptyset$; s = $\emptyset$; $t_{estart}=0$;
2: for each i $\in$[1, $|R_a|$] do
3: if $r_{ai} \leq K^*$ and $t_{estart}==0$ then
4: $t_{estart}=t_i$;
5: else if $r_{ai} > K^*$ and $t_{estart}\neq 0$ then
6: //found one event;
7: $t_{eend}=t_{i-1}$; e <$t_{estart}$,$t_{eend}$> ;
8: if $|E_a| == \emptyset$then
9: $E_a \cup = e$; $t_{sstart}=t_{sstart}$; $t_{send}=t_{send}$;
10: else if ($t_{sstart}$- $t_{send}$)< $\phi$ then
11: //e* is the last leading event before e in $E_a$;
12: $E_a \cup = e$; $t_{send}=t_{eend}$;
13: else then
14: //found one session;
15: s= <$t_{sstart}$,$t_{send}$, $E_a$> ;
16: $S_a \cup = s$; $E_a = \emptyset$; s = $\emptyset$is a new session;
17: go to Step 7;
18: $t_{estart}= 0$; e = $\emptyset$is a new leading event;
19: return $S_a$

In algorithm, e denotes leading events given in tuple as <$t_{estart}$,$t_{eend}$> and sessions are denoted as tuple <$t_{sstart}$,$t_{send}$,$E_s$> where $E_s$ is set ofleading events in leading session. Step 2 to 7 are used extract individual leading events and step 8 to 16 are used to mine leading sessions. In this way we can easily find leading events and sessions of app.

### D. Evidence Aggregation

After successful extraction of three types of evidences, the next step is combination of those evidences for ranking fraud detection. The final evidence score Ψ*(s) as a linear combination of all the existing evidences as equation given below.

## IV. CONCLUSION

In this mission, we developed up a ranking or positioning extortion discovery framework for transportable mobile Apps. In distinctive, we to begin with established that positioning misrepresentation passed off in driving periods and gave a process to digging using periods for every App from its mentioned positioning files. At that point, we famous positioning situated ranking established proofs and survey situated confirmations for detecting positioning extortion. Moreover, we proposed an enhancement founded total method to include each one of the most proofs for evaluating the validity of riding classes from the portable Apps.This paper, gives the ranking fraud detection model for mobile apps. Now a days lots of mobile app builders makes use of quite a lot of frauds strategies to broaden their rank. To avoid this, there are more than a few fraud detection techniques which are studied in this paper. We realize the ranking fraud utilizing precise fraud reports.

## REFERENCES

[1] B. Zhou, J. Pei, and Z. Tang. A spamicity approach to web spam detection. In Proceedings of the 2008 SIAM International Conference on Data Mining, SDM'08, pages 277–288, 2008.

[2] 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.

[3] 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., 2010, pp. 939–948.

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

[5] Z. Wu, J. Wu, J. Cao, and D. Tao, "HySAD: A semi-supervised hybrid shilling attack detector for trustworthy product recommendation," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 985–993.

[6] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 823–831.

[7] B. Yan and G. Chen, "AppJoy: Personalized mobile application discovery," in Proc. 9th Int. Conf. Mobile Syst., Appl., Serv., 2011, pp. 113– 126.

[8] H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian, "Exploiting enriched contextual information for mobile app classification," in Proc. 21stACMInt. Conf. Inform. Knowl. Manage., 2012, pp. 1617–1621.

[9] H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian, "Mining personal context-aware preferences for mobile users," in Proc. IEEE 12th Int. Conf. Data Mining, 2012, pp. 1212–1217.

[10] H. Zhu, H. Xiong, Y. Ge, and E. Chen, "Ranking fraud detection for mobile apps: A holistic view," in Proc.

[11] (2014) [Online]. Available: http://en.wikipedia.org/wiki/cohen's_kappa.