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 7, July 2016

A Ranking Fraud Detection System for Mobile Apps

Raghuveer Dagade, Prof. Lomesh Ahire

M.E. Student, Dept. of Computer Engineering Nutan Maharashtra Institute of Engineering & Technology, Talegaon Dabhade, Pune, India

Dept. of Computer Engineering Nutan Maharashtra Institute of Engineering & Technology, Talegaon Dabhade, Pune, India

ABSTRACT: Now a days everyone is using smart phones. There is need of various applications to be installed on smart phone. To download application smart phone user has to visit play store such as Google Play Store, Apples store etc. When user visit play store then he is able to see the various application lists. This list is built on the basis of promotion or advertisement. User doesn't have knowledge about the application (i.e. which applications are useful or useless). So user looks at the list and downloads the applications. But sometimes it happens that the downloaded application won't work or not useful. That means it is fraud in mobile application list. To avoid this fraud, we are making application in which we are going to list the applications. To list the application first we are going to find the active period of the application named as leading session. We are also investing the three types of evidences: Ranking based, Rating based and Review based evidence. Using these three evidences finally we are calculating aggregation. We evaluate our application with real world data collected form play store for long time period.

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

I. INTRODUCTION

Over the past few years the number of mobile Apps has grown at a breathtaking rate. For example, there are more than 1.6 million Apps at Google Play and Apple's App store, as of the end of July 2015. To move the development of mobile Apps, poly App stores propelled day by day App leaderboards, which show the outline rankings of most popular Apps. Truly, one of the most eventful ways for mobile Apps promoting is the App leaderboard. Countless and million dollars in income for the most part prompts a higher rank on the leaderboard. In this manner, App developers to have their Apps positioned as high as would be probable in App leaderboards for that they have a tendency to investigate distinctive courses, for example, publicizing effort to advance their Apps in such request.

However, as a recent trend, shady App engineers resort to few fake intends to intentionally support their Apps and in the end the outline rankings on an App store, rather than depending on traditional marketing solutions. This is generally executed by utilizing supposed "bot farms" or "human water armies" to filled the App downloads, ratings and reviews in a very short time.

Many mobile app stores launched day by day app leader boards which shows the chart ranking of popular apps. The leader board is the important for promoting apps. Original application grade level decreases due to the duplication arrival in the mobile apps. In recent activities duplicate version of an application not burned or blocked. This is the major defect. Higher rank leads huge number of downloads and the app developer will get more profit. In this way they allow Fake Application also. User not understanding the Fake Apps then the user also give the reviews in the fake application. Exact Review or Ratings or Ranking Percentage are not correctly Calculated. Positioning coercion in the convenient App business division insinuates misleading or beguiling practices which have an inspiration thumping up Apps in the notoriety list. Without a doubt, it turns out to be more continuous for App developers to utilize shady means, for example, blowing up their Apps business or posting imposter App appraisals, to submit positioning

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 7, July 2016

misrepresentation. While the value of anticipating positioning extortion has been for the most part seen, there is compelled understanding and examination here.

To this end, in that, give a comprehensive point of view of positioning extortion and propose a positioning misrepresentation recognition framework for portable Apps. In particular, the first propose to precisely find the mining so as to position extortion the dynamic periods, in particular driving sessions, of portable Apps. Such driving sessions can be used for perceiving the area variation from the norm instead of overall inconsistency of App rankings. Moreover, there are research three sorts of proofs, i.e. (that is) positioning based affirmations, displaying in order to rate based verifications and review based evidences, Apps' situating, rating and study practices through factual speculations tests. Moreover, a streamlining based accumulation system to join each one of the evidences for extortion discovery. At last, to evaluate the proposed framework with true App information gathered from the iOS App Store for quite a while period. In the trials, approve the viability of the proposed framework, and demonstrate the versatility of the discovery calculation and also some normality of positioning extortion exercises.

II. PROJECT IDEA

The issue of detecting ranking fraud for mobile Apps is still under-investigating. In this work, propose to add a ranking fraud detection system for mobile Apps. In that, to recognize a few imperative difficulties. Initially, ranking fraud does not generally happen in the whole life cycle of an App, so there is have to identify the exact time when fraud happens. In the local anomaly detecting such type of challenges rather than global anomaly of mobile Apps. Second, for the huge number of mobile Apps, it is hard to physically name ranking fraud for each App, so there is needed to have a versatile approach to automatically detect ranking fraud without utilizing any benchmark data. Lastly, because of the dynamic way of outline rankings, it is difficult to confirm and identify the evidences connected to ranking fraud, which moves us to find some outright misrepresentation examples of portable Apps as evidence.

To be sure, our watchful perception uncovers that mobile Apps are not generally positioned high in the leaderboard, but rather just in some leading events, which shape diverse leading sessions. At the end of the day, ranking fraud for the most part happens in these leading sessions. Hence, leading session ranking fraud is nothing but the mobile apps ranking fraud. In particular, yet powerful calculation to recognize the leading sessions of each App taking into account its historical ranking records. At that point, with the investigation of Apps positioning practices, find that the beguiling Apps as often as possible have particular positioning examples in every driving session contrasted and ordinary Apps. Along these lines, we portray some fraud evidences from Apps' historical ranking records, and create three functions to concentrates such ranking based fraud evidences. In any case, the ranking based evidences can be influences by App developers' reputation and some honest to goodness promoting effort, for example, "limited-time discount". Subsequently, it is not adequate to just utilize positioning based proofs. Therefore, further propose two types of fraud evidences based on Apps' rating and review history, which mirror some oddity designs from Apps' authentic rating and review records. Furthermore, to add to an unsupervised evidence-aggregation system to coordinate these three sorts of proofs for assessing the believability of leading sessions from mobile Apps.

III. MOTIVATION

Ranking extortion in the mobile App market implies fake or boggling exercises which have a purpose behind thumping up the Apps in the prominence list. In fact, it ends up being more progressive for App designers to use shady means, for instance, swelling their Apps' arrangements or posting faker App positioning, to submit positioning misrepresentation. While the noteworthiness of balancing Ranking extortion has been extensively seen. To give an all comprehensive perspective of ranking extortion and propose a positioning misrepresentation discovery framework for portable Apps. In particular, there is first propose to precisely find the mining so as to position extortion the dynamic periods, to be specific driving sessions, of versatile Apps. Such driving sessions can be utilized for distinguishing the neighborhood oddity rather than worldwide inconsistency of App rankings Apps is really to identify positioning misrepresentation inside of driving sessions of versatile Apps.

IV. LITERATURE SURVEY

In the work [2] Author said another system, called "Idea of fraud identification" Advances in GPS following innovation have empowered us to introduce GPS beacons in city taxis to accumulate a lot of GPS follows under

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 7, July 2016

running time limits. These GPS follows give a parallel chances to us to unwrap taxi driving misrepresentation exercises. In this work, to build up a taxi driving extortion recognition framework, which is proficient to methodically examine taxi driving misrepresentation. In this framework, first give capacities to discover two certainties of confirmations: travel course proof and driving separation proof. Besides, a third capacity is intended to consolidate the two truths of confirmations in light of dumpster Shafer hypothesis. To actualize the framework, first distinguish intriguing locales from a lot of taxi GPS logs. At that point, propose a parameter free strategy to mine the travel course confirms. Additionally, acquaint course stamp with speak to a run of the mill driving way from a fascinating site to another. In view of course check, abuse a generative measurable model to portray the conveyance of driving separation and distinguish the driving separation confirmations. At long last, assess the taxi driving extortion recognition framework with vast scale certifiable taxi GPS logs. In the trials, reveal some consistency of driving misrepresentation exercises and examine the inspiration of drivers to submit an analyzing so as to drive extortion the created taxi misrepresentation information.

In the work [3] Author said another method, called "Idea of extracting of rating and review". The plans to recognize clients creating spam surveys or audit spammers. To distinguish a few trademark practices of audit spammers and model these practices for to identify the spammers. In particular, attempt to locate the model after practices. To begin with, spammers might target specific items or item assembles keeping in mind the end goal to augment their impact. Second, they have a tendency to separate from the other commentator in their evaluations of items. The propose scoring techniques to quantify the level of spam for every commentator and apply them on an Amazon audit dataset. At that point select a sub set of profoundly uncertain commentators for further examination by our client evaluators with the assistance of an electronic spammer assessment programming uniquely created for client assessment tests. The outcomes demonstrate that our proposed positioning and directed techniques are successful in finding spammers and beat other gauge strategy taking into account accommodation votes alone. Presently, at last demonstrate that the recognized spammers have more critical effect on evaluations contrasted and the unhelpful commentators.

In the work [4] Author said a new technique, called “Concept of review analysis”. Evaluative writings on the Web have turned into a profitable wellspring of feelings on items, administrations, occasions, people, and so on. As of late, numerous scientists have concentrated such sentiment sources as item audits, discussion posts, and web journals. Be that as it may, existing exploration has been centered on grouping and rundown of assessments utilizing regular dialect handling and information mining methods. An imperative issue that has been disregarded so far is assessment spam or dependability of online suppositions. In this work, to study this issue in the setting of item audits, which are assessment rich and are generally utilized by shoppers and item producers. In the previous two years, a few new businesses additionally showed up which total assessments from item audits. It is consequently high time to study spam in surveys. To the best of our insight, there is still no distributed study on this theme, in spite of the fact that Web spam and email spam have been explored broadly. To find that sentiment spam is entirely not quite the same as Web spam and email spam, and in this manner requires diverse identification methods. In view of the examination of 5.8 million surveys and 2.14 million analysts from amazon.com, we demonstrate that conclusion spam in audits is far reaching. This work dissects such spam exercises and displays some novel procedures to recognize them.

In the work [5] Author said another strategy, called "Concept of Rank aggregation". Numerous applications in data recovery, common dialect handling, information mining, and related fields require a positioning of occurrences concerning indicated criteria rather than a characterization. Moreover, for some such issues, numerous built up positioning models have been all around examined and it is attractive to consolidate their outcomes into a joint positioning, formalism signified as rank conglomeration. This work displays a novel unsupervised learning calculation for rank accumulation (ULARA) which gives back a straight mix of the individual positioning capacities taking into account the guideline of compensating requesting assertion between the rankers. Notwithstanding displaying ULARA, we show its adequacy on an information combination assignment crosswise over specially appointed recovery frameworks.

V. IMPLEMENTATION STRATEGIES

To proposed an enhancement based aggregation technique to coordinate every one of the evidences for assessing the validity of leading sessions from mobile Apps. An extraordinary viewpoint of this methodology is that every one of the evidences can be modeled by statistical hypothesis tests, along these lines it is anything but difficult to be reached out with different proofs from area information to recognize positioning extortion. At long last, approve the proposed framework with broad investigations on genuine App information gathered from the Apple's App store. To identified

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 7, July 2016

- Ranking Based Evidence
- Rating Based Evidence
- Review Based Evidence

In Ranking Based Evidences, by examining the Apps' authentic positioning records, we watch that Apps' positioning practices in a main occasion dependably fulfill a particular positioning example, which comprises of three diverse positioning stages, in particular, Rising phase, maintaining phase and recession phase.

In Rating Based Evidences, particularly, after an App has been distributed, it can be evaluated by any client who downloaded it. To be sure, client rating is a standout amongst the most imperative components of App commercial. An App which has higher rating might draw in more clients to download and can likewise be positioned higher in the pioneer board. Consequently, evaluating control is additionally an essential point of view of positioning misrepresentation.

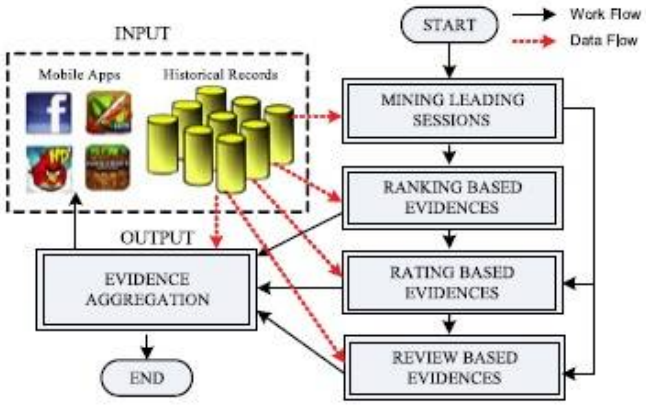

Fig: The framework of ranking fraud detection system for mobile Apps.

In Review Based Evidences, other than evaluations, the vast majority of the App stores additionally permit clients to compose some literary remarks as App surveys. Such review can mirror the individual recognitions and use encounters of existing clients for specific mobile Apps. In fact, review control is a standout amongst the most critical viewpoint of App positioning misrepresentation.

Finally, when all evidences extracted then aggregate all evidences and provide single rank. Which is genuine rank free from fraud. Here uses root mean square error testing used to test the evidences.

This work use mining leading session algorithm for extracting all evidences of raking fraud in mobile apps. In that, first we identify the leading event of all mobile app's which is available in leaderboard. After that finding the leading session of each individual app's for the purpose of identifying the genuine evidences. Today the dynamic nature of giving rating ranking and review to any apps. So it is not possible to tag each individual app its fraud or not manually. There require automatic system to analyze the apps evidences.

To provides constrain to the user to giving ranking, rating and review to the mobile apps. Because more than 1.6 billon of are available in paly store. So, it is difficult to analyze the evidences of fraud apps. When we use constrain to user then user can only give the genuine ranking, rating and review to apps.

In this work also provide the app recommendation system for mobile apps. When user searching app for downloading there user can able to see different related apps on screen. KNN (K- Nearest Neighbor) algorithm is use to finding the related apps from the different classes. KNN algorithm used for app recommendation system.

This algorithm is based on the observation that a sample that has features that are similar to the ones of points of one particular class it belongs to that class. These points are known as nearest neighbors. The parameter k specifies the number of neighbors (neighboring points) used to classify one particular sample point.

A percentage of the critical properties of the KNN algorithm are listed below:

- The KNN can be utilized to characterize information without requiring model building, this is called "instance based learning".
- An estimation of the separation between information focuses ought to be accessible.

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 7, July 2016

- The KNN order is construct exclusively in light of nearby data (just the k closest neighboring information focuses are examined amid the arrangement process).
- The decision boundaries produced by the KNN classification can of arbitrary shape.
- The grouping is delicate to the right determination of k. In the event that k is too little it will prompt "over-fitting".

VI. MATHEMATICAL MODELING

Set Theory Analysis

Our system can be represented as a set

1) *System S = {I, O, C}*

Where,

I=set of inputs
O=set of outputs
C = set of constraints

2) *Input*
Input I = {Login, Request}
Login = {Username, Password}
Request = {Search apps, search top apps, download app, Apply rating and review, Find Fraud, List apps, View History}
Users = {User, Service provider}
Username = {Username1, Username2… Username n}
Password = {Password$_1$, Password$_2$… password $_n$}

3) *Output*
Output O = {Display fraud in apps, Download start, display app list, Display history}

4) *Constraint*
C = "User should login to the system before its usage".

5) *Space Complexity: The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.*

6) *Time Complexity: Check No. of patterns available in the datasets= n, If (n>1) then retrieving of information can be time consuming. So the time complexity of this algorithm is* $O(n^n)$.

Φ = *Failures and Success conditions.*

7) *Failures:*
1. Huge database can lead to more time consumption to get the information.
2. Hardware failure.
3. Software failure.

8) *Success:*
1. Search the required information from available in Datasets.
2. User gets recommendation for products.

VII. EXISTING APPROACH

- Due to the dynamic way of outline rankings, it is difficult to recognize and affirm the confirmations connected to positioning misrepresentation, which inspires us to find some certain extortion examples of portable Apps as proofs.
- Cannot ready to identify ranking fraud happened in Apps' authentic (historical) leading session.
- There is no current benchmark to choose which leading sessions or Apps truly contain ranking extortion.
- Although a percentage of the current methodologies can be utilized for inconsistency identification from verifiable rating and review records, they are not ready to concentrate extortion confirmations for a given time period (i.e., leading session).

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 7, July 2016

VIII. PROPOSED APPROACH

- The proposed structure is adaptable and can be stretched out with other space produced confirmations for positioning extortion identification.
- Experimental results demonstrate the adequacy of the proposed framework, the versatility of the identification calculation and in addition some normality of positioning misrepresentation exercises.
- To the best of our insight, there is no current benchmark to choose which driving sessions or Apps truly contain positioning extortion. In this manner, we create four natural baselines and welcome five human evaluators to accept the viability of our methodology Evidence Aggregation based Ranking Fraud Detection (EA-RFD).
- After coming to and keeping up the normal positioning for a required period, the control will be ceased and the positioning of the malignant App will diminish significantly.
- The review based confirmations could simply enhance the identification exhibitions while being utilized together with different proofs.

IX. OUTCOMES

- App provider uploads the apps in our system.
- First to analyses the fraud in uploaded apps we have to collect historical records of apps.
- After collecting records we have to extract that records.
- From this extraction process we have to find mining leading sessions, ranking evidences, rating evidences and review evidences.
- After extracting all evidences we have to perform aggregation on all evidences.
- Using that aggregation we have to provide ranks for that apps.

This will eliminates the fraud in ranking of mobile apps.

X. GOALS AND OBJECTIVES

For this work, plan to concentrate more powerful misrepresentation proves and examine the inactive relationship among rating, review and rankings. Additionally, to amplify our ranking fraud discovery approach with other mobile App related administrations, for example, mobile Apps suggestion, for upgrading client experience. The future works are about giving the security to mobile applications much proficient way. This works can likewise include imperatives client for giving review, ranking and rating furthermore downloading applications.

- To rank fraud for mobile application.
- To enhance the fraud detection efficiency.
- First examinations the fundamental qualities of leading events for separating misrepresentation confirmations.
- The suspicious leading events might contain short rising and recession phases.
- Analyses web ranking spam detection. In particular, the web ranking spam alludes to any intentional activities which convey to choose site pages a baseless ideal significance or significance.
- Focused on recognizing online review spam.

XI. CONCLUSION

In this work, to added to a positioning extortion discovery framework for versatile Apps. In particular, we initially demonstrated that ranking fraud happened in leading sessions and gave a strategy to mining leading sessions for each App from its historical records. At that point, we distinguished ranking based evidences, rating based proofs and review based confirmations for recognizing ranking fraud. In addition, we proposed an advancement based total technique to coordinate every one of the confirmations for assessing the validity of leading sessions from mobile Apps. A one of a kind viewpoint of this methodology is that every one of the confirmations can be modeled by statistical hypothesis tests, therefore it is anything but difficult to be stretched out with different evidences from domain knowledge to detect ranking fraud.

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 7, July 2016

ACKNOWLEDGMENT

I might want to take this chance to express my significant appreciation and profound respect to my Guide Prof. Lomesh Ahire, for her excellent direction, valuable feedback and steady support all through the length of time of the paper. The valuable proposals from the aide were of huge help all through my paper. Her insightful feedback kept me attempting to make this paper in a much better way. Working under her was an extremely knowledgeable experience for me.

REFERENCES

[1] Hengshu Zhu, Hui Xiong, Yong Ge, Enhong Chen, "Discovery of Ranking fraud for Mobile Apps," in Proc. IEEE Tran. On Knowledge and Data Engineering, vol.27, No.1, January 2015..
[2] 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.
[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] N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219-230.
[5] 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.
[6] A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with distance-based models," in Proc. 25th Int. Conf. Mach. Learn., 2008, pp. 472–479.
[7] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in Proc. 15th Int. Conf. World Wide Web, 2006, pp. 83-92.K.
[8] N. Spirin and J. Han, "Survey on web spam detection: Principles and algorithms," SIGKDD Explor. Newslett., vol. 13, no. 2, pp. 50-64, May 2012.
[9] 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.
[10] 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.
[11] 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.
[12] B. Yan and G. Chen, "AppJoy: Personalized mobile application discovery," in Proc. 9th Int. Conf. Mobile Syst., Appl., Serv., 2011, pp. 113-126.
[13] H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian, "Exploiting enriched contextual information for mobile app classification," in Proc. 21st ACMInt. Conf. Inform. Knowl. Manage., 2012, pp. 1617-1621.
[14] 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.
[15] D. F. Gleich and L.-h. Lim, "Rank aggregation via nuclea norm minimization," in Proc.17th ACM SIGKDD Int Conf. Knowl. Discovery Data Mining, 2011, pp. 60-68.