International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 4, Issue 2, No 4, 2016. ISSN: 2348-6600
Reference ID: IJCS-131 PAGE NO: 772-777

Mobile App Recommendation & Ranking Fraud Detection on Relationship among Rating Review & Ranking

E.Ramya
Research Scholar, Department of Computer Science
Shrimati Indira Gandhi College, Trichy,
Mail id: Ramyaelango93@gmail.com
Mrs.V.Vetriselvi, M.sc.,Mphil.,
Assistant Professor,Department of Computer Science,
MCA and Information Technology and Application, Shrimati Indira Gandhi college, Trichy
Mail id: vetriselviramesh@gmail.com

***Abstract*—** Mobile application plays an important role for all the smart phone users to play or perform different tasks. Mobile application developers are available in large number; they can develop the different mobile applications. For making lager users for their applications some developers involve in illegal activities. Due to these illegal activities the mobile applications hires high rank in the application popularity list. Such fraudulent activities are used by more and more application developers. A ranking fraud detection system for mobile Apps is proposed in this paper. Accurately locate the ranking fraud by mining the leading sessions, of mobile Apps.R3-RFD algorithm is proposed in this paper. Furthermore, sentiword dictionary is used to identify the exact reviews scores. The fake feedbacks by a same person for pushing up that app on the leaderboard are restricted. Two different constraints are considered for accepting the feedback given to an application. The first constraint is that an app can be rated only once from a user login. And the second is implemented with the aid of MAC address that limits the number of user login logged per day from a MAC address as five.

Keywords: *Mobile Applications, fraud detection, evidences, Historical Record*

1. INTRODUCTION

Ranking fraud in the mobile app market refers to fraudulent or deceptive activities which have a purpose of bumping up the apps in the popularity app developers to use shady means, such as inflating their apps ' sales or pos ting phony App ratings , to commit ranking fraud. While the importance of preventing ranking fraud has been widely recognized, there is limited understanding and research in this area. To this end, in this paper, we provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile apps. Specifically, we first propose to accurately locate the ranking fraud by mining the active periods, namely leading sessions, of mobile Apps. Such leading sessions can be leveraged for detecting the local anomaly instead of global anomaly of app rankings. Furthermore, we investigate three types of evidences, i.e., ranking based evidences, rating based evidences and review based evidences, by modeling apps ' ranking, rating and review behaviors through statistical hypotheses tests. In

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 4, Issue 2, No 4, 2016. ISSN: 2348-6600
Reference ID: IJCS-131 PAGE NO: 772-777

addition, we propose an optimization based aggregation method to integrate all the evidences for fraud detection.

Purpose

The purpose of this project is to provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile apps. Ranking fraud in the mobile app market refers to fraudulent or deceptive activities which have a purpose of bumping up the apps in the popularity list. Specifically, first propose to accurately locate the ranking fraud by mining the active periods, namely leading sessions, of mobile Apps. Three types of evidences, i.e., ranking based evidences, rating based evidences and review based evidences are used.

2. LITERATURE SURVEY

1.D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation" [1]:

D. M. Blei, A. Y. Ng, and M. I. Jordan, introduces a unique model called as Dirichlet allocation (LDA) a generative probabilistic model for collections of discrete data such as text amount. Basically it is a three level hierarchical Bayesian model in which each element of a group is demonstrated as a finite mixture over a fundamental set of topics. Each topic is demonstrated as an infinite mixture over fundamental set of topic probabilities. With the reference of text modelling, the topic probabilities provide an open representation of a document. An efficient approximation inference technique is presented based on various methods and an EM algorithm for empirical Bayes parameter estimation is also presented. The results are reported in document modelling, text classification and collaborative filtering, which compares to a collection of unigrams and probabilistic LSI model.

2.Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "A taxi driving fraud detection systemices in city taxis"[2]:

Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou , illustrated that growth in the field of GPS tracking technology have allowed the users to install GPS tracking devices in taxies to gather huge amount of GPS traces under some time period. These traces by GPS offered an unparalleled opportunity to uncover taxi driving fraud traces and then fraud detection system is proposed which is able to identify taxi driving fraud. First, two sort of function are uncovered here i.e. travel route evidence and driving distance evidence. Even a third function is developed to combine the previous functions based on Dempster-Shafer theory. First identification of interesting locations is done from tremendous amount of taxi GPS logs and then parameter free method is proposed to extract the travel route evidences. Secondly, concept of route mark is
+++developed to illustrate the driving path between locations and based on those mark, specific model is characterized for the distribution of driving distance and discover the driving distance evidences. And finally, taxi driving fraud detection system with a large scale real world taxi GPS logs.

3. T. L. Griffiths and M. Steyvers, "Rank aggregation via nuclear norm minimization" [3]:

T. L. Griffiths and M. Steyvers, introduces the process of rank aggregation which is interweave with the structure of skew-symmetric matrices. Recent development in the principles of matrix completion matrices is been applied and this idea gives rise to a new method for ranking a set of items. The core of this idea deals with the raking aggregation method which intimately describes a

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 4, Issue 2, No 4, 2016. ISSN: 2348-6600
Reference ID: IJCS-131 PAGE NO: 772-777

partially filled skew-symmetric matrix. The algorithm for matrix completion is raised to hold skew-symmetric data and use that to extract ranks for each item. This algorithm applies same strategy for both pairwise comparisons as well as for rating data. It becomes robust to noise and incomplete data as it is based on matrix completion.

4. A. Klementiev, D. Roth, and K. Small , "An unsupervised learning algorithm for rank aggregation"[6]:

A. Klementiev, D. Roth, and K. Small, describes the field of information retrieval, data mining, and natural language, many applications needs a ranking of instances which is not present in classification. Furthermore, a rank aggregation is a result of aggregating the results of the established ranking models into a formalism and then result represents a novel unsupervised learning algorithm(ULARA)which gives a linear combination of individual ranking functions. These functions were developed based on the axiom of rewarding ordering agreement between the rankers.

5. A. Klementiev, D. Roth, and K. Small, "**Unsupervised rank aggregation with distance-based models" [8]:**

A. Klementiev, D. Roth, and K. Small, produces a model which has to integrate the set of rankings often deals with aggregating and it only comes up when a certain ranked data is developed. Even though the various heuristic and supervised learning approaches to rank aggregation, a requirement of domain knowledge and supervised ranked data
exists. Therefore, to solve this issue, a framework is proposed for learning aggregate rankings without supervision. This framework is instantiated for the cases of permutations and combinations of top-k lists.

3. Existing System

The number of mobile Apps has grown at a breathtaking rate over the past few years. For example, as of the end of April 2013, there are more than 1.6 million Apps at Apple's App store and Google Play. To stimulate the development of mobile Apps, many App stores launched daily App leaderboards, which demonstrate the chart rankings of most popular Apps. Indeed, the App leaderboard is one of the most important ways for promoting mobile Apps. A higher rank on the leaderboard usually leads to a huge number of downloads and million dollars in revenue. Therefore, App developers tend to explore various ways such as advertising campaigns to promote their Apps in order to have their Apps ranked as high as possible in such App leaderboards.

Drawbacks of Existing System

- The problem of detecting ranking fraud for mobile Apps is still under-explored
- Due to the huge number of mobile Apps, it is difficult to manually label ranking fraud for each App.

Proposed System

In this paper, we propose to develop a ranking fraud detection system for mobile Apps. Along this line, we identify several important challenges. First, ranking fraud does not always happen in the whole life cycle of an App, so we need to detect the time when fraud happens. Such challenge can be regarded as detecting the local anomaly instead of global anomaly of mobile Apps. Second, due to the huge number of mobile Apps, it is difficult to manually label ranking fraud for each App, so it is important to have a scalable way to automatically detect ranking fraud without using any benchmark information. Finally, due to the dynamic nature of chart rankings, it is not easy to identify and confirm the evidences linked to

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 4, Issue 2, No 4, 2016. ISSN: 2348-6600
Reference ID: IJCS-131 PAGE NO: 772-777

ranking fraud, which motivates us to discover some implicit fraud patterns of mobile Apps as evidences.

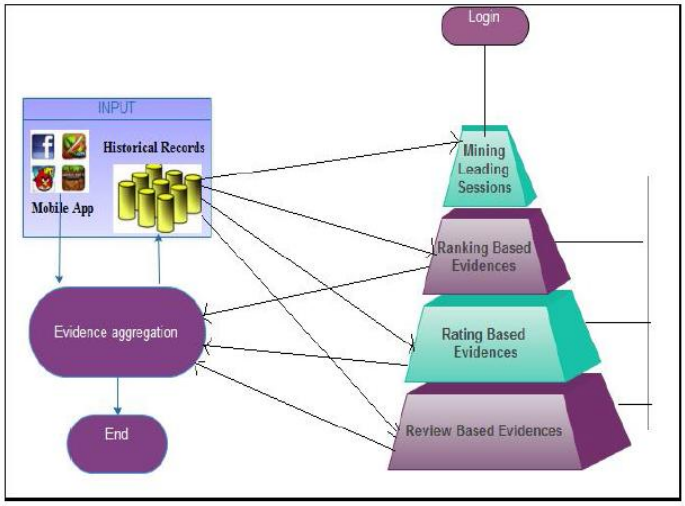
Detecting ranking fraud of mobile Apps is actually to detect ranking fraud within leading sessions of mobile Apps. Specifically, we first propose a simple yet effective algorithm to identify the leading sessions of each App based on its historical ranking records. Then, with the analysis of Apps' ranking behaviors, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, we characterize some fraud evidences from Apps' historical ranking records, and develop three functions to extract such ranking based fraud evidences.

4.Problem Statement

Many mobile app stores launched daily app leader boards which show 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 they Allow Fake Application also. User not understanding the Fake Apps then the user also gives the reviews in the fake application. Exact Review or Ratings or Ranking Percentage are not correctly Calculated.

5. SYSTEM MODEL

In recent years, mobile app has been rapidly growing while boosting more than 400,000 applications like Apple app store and Google Android market. This rapid improvement of mobile App has made it complex to user for finding unique and trusted patterns of Application. Thus to solve this issue, marketing executives use ranking for the App. In this paper, a useful R3-RFD algorithm is used to find the leading sessions and with the analysis of those records, it is proved that apps usually have different ranking patterns in each sessions as compared to the normal apps. Therefore it is illustrated from those ranking records that some fraud is taking place in mobile app market and to restrict those frauds, evidences are developed to detect such fraud. As only ranking based evidences does not seems to be much sufficient to detect the fraud of mobile app, based on apps rating and review history some fraud evidences were discovered which showed anomaly patterns by those history. Specifically, an unsupervised evidence aggregation method is also proposed for evaluating the trustworthiness of leading sessions. And finally, the proposed system is estimated with real world app data gathered from the Google Play store for time consuming period. The results of these experiments showed an effectiveness of proposed approach in fig



International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 4, Issue 2, No 4, 2016. ISSN: 2348-6600
Reference ID: IJCS-131 PAGE NO: 772-777

6.Experimental Results

Experimental Data

- The experimental data sets were collected from the "Top Free 300" and "Top Paid 300" leaderboards of Google's Play Store.
- The data sets contain the daily chart rankings1 of top 300 free Apps and top 300 paid Apps, respectively. Moreover, each data set also contains the user ratings and review information.

Mining Leading Sessions

Set the ranking threshold K*=300 and threshold φ= 7

There are two main steps for mining leading sessions. First, we need to discover leading events from the Apps historical ranking records. Second, we need to merge adjacent leading events for constructing leading sessions. Specifically, Algorithm 1 demonstrates the pseudo code of mining leading sessions for a given App a.

Algorithm 1 Mining Leading Sessions

Input 1: a's historical ranking records R_a;
Input 2: the ranking threshold K^*;
Input 2: the merging threshold ϕ;
Output: the set of a's leading sessions S_a;
Initialization: $S_a = \emptyset$;

$E_s = \emptyset$; $e = \emptyset$; $s = \emptyset$; $t^e_{start} = 0$;
for each $i \in [1, |R_a|]$ **do**
  if $r^a_i \leq K^*$ **and** $t^e_{start} == 0$ **then**
    $t^e_{start} = t_i$;
  else if $r^a_i > K^*$ **and** $t^e_{start} \neq 0$ **then**
    //found one event;
    $t^e_{end} = t_{i-1}$; $e =< t^e_{start}, t^e_{end} >$;
    if $E_s == \emptyset$ **then**
      $E_s \cup = e$; $t^s_{start} = t^e_{start}$; $t^s_{end} = t^e_{end}$;
    else if $(t^e_{start} - t^s_{end}) < \phi$ **then**
      $E_s \cup = e$; $t^s_{end} = t^e_{end}$;
    else then
      //found one session;
      $s =< t^s_{start}, t^s_{end}, E_s >$;
      $S_a \cup = s$; $s = \emptyset$ is a new session;
      $E_s = \{e\}$; $t^s_{start} = t^e_{start}$; $t^s_{end} = t^e_{end}$;
    $t^e_{start} = 0$; $e = \emptyset$ is a new leading event;
return S_a

6. CONCLUSION AND FUTURE WORK

This paper introduces a system which is built up and it is actually a positioning extortion discovery framework for mobile Apps. In particular, initially it is demonstrated that positioning misrepresentation happened in driving sessions and gave a system to digging driving sessions for each App from its chronicled positioning records. At that point, it is recognized that positioning based confirmations, rating based proofs and survey based confirmations are used for identifying positioning extortion. In addition, a unique model is proposed which is an improvement based total system to incorporate every one of the proofs for assessing the validity of driving sessions from portable Apps. A novel point of view of this methodology is that every one of the proofs can be displayed by measurable theory test, in this way it is anything but difficult to be reached out with different confirmations from space information to distinguish positioning misrepresentation. At last, the proposed framework is accepted with broad examinations on certifiable App information gathered from the Apple's App store. Exploratory results demonstrated the adequacy of the proposed methodology. Later on, to concentrate more viable misrepresentation confirms and dissect the idle relationship among rating, survey and rankings is planned. In addition, amplification of positioning misrepresentation location approach is performed with other portable App related administrations, for example, mobile Apps suggestion, for improving client experience.

8. REFERENCES:

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., pp. 993–1022, 2003.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 4, Issue 2, No 4, 2016. ISSN: 2348-6600
Reference ID: IJCS-131 PAGE NO: 772-777

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

AUTHORS

First Author – E.Ramya
Research Scholar, Department of Computer Science,Shrimati Indira Gandhi College, Trichy,
Mail id: Ramyaelango93@gmail.com

Second Author – Mrs.V.Vetriselvi, M.sc.,Mphil.,
Assistant Professor,Department of Computer Science,MCA and Information Technology and Application, Shrimati Indira Gandhi college, Trichy
Mail id: vetriselviramesh@gmail.com