Integrated Approach for Mobile App Ranking Fraud Detection

Aadil Khan¹ & Prof. T. H. Gurav²
¹,²Savitribai Phule Pune University, Smt. Kashibai Navale College of Engineering, Pune, India.

Abstract: Ranking fraud in Applications for mobiles propose to false or precarious exercises which have an intent in background, to push the Apps high in the rank list. In recent days, various shady means are used more much of the time by application developers, such developing their Apps’ business or posting fraud App evaluations, to give situating contortion. There is a limited comprehension and examination region of study to prevent such ranking frauds. This paper gives holistic view of ranking manipulations and portrays a Ranking fraud identifiable framework for mobile Apps. This work is done into three classifications. At first, ranking fraud discovery second is online review identification and last one is mobile application suggestions. The Web ranking demonstrates to any purposeful activities which pass on to choose Web pages an unmerited ideal pertinence or importance. Survey is expected to give unfair for perspective of a couple of items to affect the clients’ perspective of the items by particularly or by implication blowing up or harming the item’s notoriety. Proposed framework additionally eliminates the fake surveys from the dataset utilizing same measure algorithm and after that distinguish the application rank. At last this system will also recommend Apps which are more relevant and most genuine. The propose framework will saves the time and also memory than the previous framework.

Keywords-- Mobile Apps, Ranking Fraud Detection, Evidence Aggregation, Historical Ranking Records, Rating and Review, Recommendation.

1. Introduction

AS smartphones emerges new technologies like android and iOS operating system took a boost in market. Mobile application started growing at such a high rate. As a study says millions of apps are there on apple’s app store and on Google Play. This started a new business in computer world and became a reason to earn thousands of dollars and downloads. Daily leaderboard is published by these markets contains the most popular apps which will consequently be downloaded and rated most high by users. Some developers may use some marketing strategies like an advertisement campaign for promotion of their app.

However this part of technology is also not safe from threats. Mobile app market, we refer it as market, is manipulated by some fraudulent app developers to bump up their app high in the rank list, as an app in leaderboard confirms high downloads and high income. Shady means are used to make such a fraud and implemented using “bot farms” which is also called “Human water armies”.

In this area some related work is there, for example, spam detection for web ranking, mobile app recommendations, and some online review based spam detection. Our study thus focuses on an integrated approach, for various evidences, to find Mobile App ranking fraud and also recommend the most relevant App that is most genuine. For this we have to go through challenges like first we need to find at what time the fraud is happening it means exact time of fraud is needed. Secondly we know that there are tremendous number of Apps present in market so it is nearly impossible to physically mark ranking fraud for every App, so it’s crucial to automatically distinguish fraud without utilizing any essential data.

Mobile Apps are not commonly ranked high in the leader board, but instead just in a few events ranking frauds more often than not happens in leading sessions. In this way, fundamental target is to recognize ranking fraud of mobile Apps inside of leading sessions. Initially propose an efficient algorithm to recognize the main sessions of every App depends on its previous ranking records. By then, with the examination of Apps’ ranking practices, find the fake Apps consistently have unique ranking examples in every leading session contrasted with ordinary Apps. Along these lines, some fraud confirmations are portrayed from Apps’ previous ranking records. By then three limits are produced to concentrate such ranking based fraud confirmations. Thusly, help two kinds of fraud confirmations are proposed taking into account Apps’ rating and survey history, which mirror some
Inconsistency patterns from Apps’ previous rating and review records. Also, to coordinate these three kinds of unsupervised proof collection procedure is created which is used for assessing the validity of leading sessions from mobile Apps.

In this paper study about the related work done, in section II, the implementation details in section III where see the system architecture, modules description, mathematical models, algorithms and experimental setup. In section IV discuss about the expected results and at last provide a conclusion in section V.

2. Related Work.

In this section discuss existing work done by the researchers for text mining process.

In paper [1], author made ranking fraud location framework for mobile Apps. Specifically, they at first showed that ranking fraud happened in leading sessions and gave a strategy to mining leading sessions for every App from its previous ranking records. By then, they recognized ranking based confirmations, rating based proofs and review based proofs for distinguishing ranking fraud. They likewise proposed a optimization based collection strategy to consolidate each one of the proofs for surveying the validity of leading sessions from mobile Apps.

In paper [2], author have focused on various parts of substance build spam regarding the Web and showed different heuristic schedules for recognizing content based spam. Here, they continue with examinations of "web spam": the infusion of misleadingly made pages into the web with a particular deciding objective to affect the results from web crawlers, to direct individuals to particular pages for the purpose of excitement or advantage.

In paper [3], author has reported a review on Web spam location, which altogether exhibits the rules and algorithm in the literature. Undoubtedly, the work of Web ranking spam identification is principally in light of the examination of ranking measures of web searchers, for instance, PageRank and question term frequency. This is not the same as ranking fraud location for mobile Apps. They sort each present algorithm into three classifications in light of the type of information they use: content-based techniques, link based strategies and techniques on the basis of non-conventional information, for instance, customer behavior, snaps, and HTTP sessions.

In paper [4], authors have seen a couple of representative behaviors of review spammers and model these practices to identify the spammers. This paper expects to perceive clients creating spam overviews or review spammers. They perceive a couple trademark practices of review spammers and model these practices with a specific end goal to recognize the spammers. Authors attempt to exhibit the going with practices. In any case, spammers may target specific things or item stores up in order to grow their impact. Second, they have a tendency to go out of order from exchange experts in their assessments of items.

In paper [5], authors have analyzed the issue of discovering hybrid shilling attacks on rating information. The methodology depends on can be utilized for reliable item suggestion and the semi-supervised learning. This paper displays a Hybrid Shilling Attack Detector or HySAD for short, to handle this issue. Specifically, HySAD familiarizes MC-Relief with select successful recognition metrics and Semi managed Naive Bays (SNB) to correctly isolate Random-Filler model aggressors and Average-Filler model attackers from standard clients.

In paper [6], authors have analyzed the issue of singleton survey spam detection. Specifically, they handled this issue by recognizing the co-anomaly pattern in different review based time arrangement. Also some of above strategies can be used for anomaly detection from previous rating and overview records, they are not prepared to focus fraud evidences for a given time period (i.e., leading session).

In paper [7], author created a mobile App recommender framework, Appjoy, which depends on user’s App use records to assemble an inclination matrix despite utilizing explicit client ratings.

In paper [8], author analyzed a few suggestion models and proposed a content-based collaborative separating model, called Eigenapp, for prescribing Apps in their Web website Getjar. Also, a few researchers analyzed the issue of misusing advanced logical data for mobile App suggestion.

3. Implementation Details

3.1 System Overview

In the present framework proposes a ranking fraud detection system for mobile Applications. Ranking fraud does not create in the life cycle of particular Apps, so this need to perceive when the fraud occurred. According to the proposed framework experimentation it demonstrates the mobile Apps are not commonly ranked high in the leaderboard, in spite of the leading events, which shape particular leading sessions. Like this ranking fraud happens in the leading sessions. Particularly this framework first
proposes an essential successful algorithm to recognize the leading session of every Application depends on previous records. At that point with the exploration of Apps ranking behaviors, this framework analyzes the fake Apps as often as possible have different ranking patterns in each leading session observed with ordinary Apps. The primary contribution of this framework is to find the fake reviews and eliminates it.

Where,

Module 1: Remove Fake Reviews
By taking input data which contains reviews and ratings, this system identifies the replicated and fake reviews and removes that reviews. For this system it use J48 classification algorithm which create testing file. The proposed system compares it with reviews and identifies the fake reviews and removes it. It preserves the time and also more memory while operation is going on.

Module 2: Leading events
Given a ranking limit a basic event e of Apps a contains a period range additionally, relating rankings of a. Note that ranking edge K * is used which is specifically more than K here in light of the way that K may be huge (e.g., more than 1,000) and the ranking records past K (e.g., 300) are not incredibly accommodating for identifying the ranking controls. Additionally, it is finding that a couple of Applications have a couple of close-by leading even which are almost each other and structure a lead session.

Module 3: Leading Sessions
Fundamentally the primary sessions of mobile application mean the time of predominance, thus these leading sessions will incorporate ranking control only. Thusly, the issue of distinguishing ranking fraud is to identify tricky leading sessions. Nearby the crucial work is to remove the leading sessions of a mobile App from its previous ranking records.

Module 4: Identifying the leading sessions for mobile apps
Basically, mining leading sessions has two kinds of steps worried with mobile fraud applications. Firstly, from the Apps previous ranking records, disclosure of leading events is done and after that combine of contiguous leading events is done which appeared for building leading sessions. Unquestionably, some specific algorithm is appeared from the pseudo code of mining sessions of given mobile application and that algorithm can recognize the particular leading events and sessions by observing the previous records one by one.

Module 5: Identifying evidences for ranking fraud detection
1) Ranking based evidences:
   It includes that leading session contains diverse leading events. Subsequently by examination of crucial behavior of leading events for finding fraud evidences besides for the application previous ranking records, it is been analyzed that a specific ranking pattern is always satisfied by application ranking behavior in a leading event.

2) Rating based evidences:
   Previous ranking based evidences are useful for distinguishing intention however it is not adequate. Considering the issue of "restrict time reduction", identifiable evidence of fraud proofs is planned due to application previous rating records. As this framework understand that rating is been done in after downloading it by the user and if the rating is high in leaderboard fundamentally that is attracted by a large portion of the mobile application users. All of a sudden, the ratings within leading session offers rise to the anomaly pattern which happens within rating fraud. These previous records can be used for making rating based proofs.

Recommendation System
The proposed system contributes the new concept of recommendation system for the mobile applications to the number of users. This is implements the apriori algorithm for the recommendations of the various applications that restrict some fake reviews for applications. The recommendation system works on the number of reviews and ratings are given by the users for the specific product. The majority of existing approaches to recommender systems target recommending the foremost relevant content to users...
improper discourse data and don’t take into consideration the danger of distressing the user in specific state of affairs. However, in several applications, like recommending personalized content, it’s additionally necessary to think about the danger of displeasing the user therefore as to not push recommendations in sure circumstances, for example, throughout knowledgeable meeting, early morning, and late night. Therefore, the performance of the recommender system depends partially on the degree to that it’s incorporated the danger into the advice method.

3.2 Algorithm

Algorithm 1 Proposed Algorithm

Input: reviews
Output: find true or fake review

1. From every analysis develop testing dataset
2. apply J48 classification algorithm on testing dataset
3. J48 classification algorithm classify testing dataset
4. on the basis of result classify fake or true review

Algorithm 2 FP-Growth

Input: MIMIC Data Set
Output: Rules
Process:
1. Data set
2. FP-growth
3. FP-tree
4. Header Table
5. Conditional FP-tree
6. Repeat step 2 to 5
7. Generate Rules

3.3 Mathematical Model

System S is represented as S= {M, H, F, L, R, T, E, A}
Input
Mobile Apps
M= {m1, m2, m3, .....mn}
Where, M is the set of mobile apps and m1, m2, m3, .....mn are the number of apps.
Historical Records
H= {h1, h2, h3, .....hn}
Where, H is represent as a set of historical records and h1, h2, h3, .....hn number of records.
Process
Remove Fake Reviews
F= {f1, f2, f3, .....fn}
Where, F is the set of fake reviews and f1, f2, f3, .....fn are the number of reviews.
Mining Leading Sessions
P= {P1, P2, P3, .....Pn}
Where, P is represent as a set of mining leading sessions and P1, P2, P3,......, In number of mining leading sessions.
Ranking Based Evidences
X= {X1, X2, X3, .....Xn}
Where, X is represent as a set of ranking based evidences and X1, X2, X3, .....Xn number of ranking based evidences.
Rating Based Evidences
Z= {Z1, Z2, Z3,.......Zn}
Where, Z is represent as a set of rating based evidences and Z1, Z2, Z3,......, Zn number of rating based evidences.
Review Based Evidences
D= {D1, D2, D3, ......, Dn}
Where, D is represent as a set of review based evidences and D1, D2, D3,......, Dn number of reviews.
Output Evidence Aggregations
B= { B1, B2, B3, .....Bn }
Where B is the set of evidence aggregations and B1, B2, B3, .....Bn represent as a number of aggregations.

3.4 Experimental Setup

This system is developed on Java framework (version jdk 6) and Netbeans (version 6.9) is used as a development tool on Windows platform. This system is able to run on any common machine as well as it doesn’t require any specific hardware.

4. Result and Dataset

4.1 Dataset

In this system use of mobile apps and historical records as dataset.

4.2 Results

<table>
<thead>
<tr>
<th>List size</th>
<th>Recommendation accuracy before fraud detection</th>
<th>Recommendation accuracy after fraud detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>68</td>
<td>78</td>
</tr>
<tr>
<td>2</td>
<td>72</td>
<td>85</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>67</td>
</tr>
<tr>
<td>4</td>
<td>67</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>78</td>
<td>89</td>
</tr>
</tbody>
</table>
5. Conclusion

In this paper, we analyzed ranking fraud detection model for mobile applications. Currently a large number of mobile application engineers use distinctive fraud frameworks to create their rank. To prevent this, there are distinctive fraud identifying techniques which are introduced in this paper. Such systems are collected into three classes, for instance, web ranking fraud recognition, online review fraud discovery, mobile application recommendation. Each one of these techniques is feasibly handling ranking fraud detection. Besides, it optimized based aggregation technique to integrate all the evidences for assessing the believability of leading sessions from mobile Apps. The recommendation system works for the mobile application recommendation system. The proposed system implements the FP-growth algorithm that work rule generation for the recommendation system that restricts the fake reviews. The system recommendation has been generated through the system FP-growth operations for the better results to the user on the basis of previous records. Our proposed system also eliminates the fake reviews from the dataset utilizing similarity measure algorithm and detect the web rank. The proposed system is saves the time as well as memory than the previous system.

6. Acknowledgement

Wherever Times is specified, Times Roman, or New Times Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times that you have access to. Please avoid using bitmapped fonts if possible. True-Type 1 fonts are preferred.

7. Reference


