A Novel Recommendation System for Mobile Apps Based on Detection of Ranking Frauds (Overview)

Pratiksha N.Magar\textsuperscript{1}, Shweta S. Dhodare\textsuperscript{2}, Poonam A.Falley\textsuperscript{3}, Priyanka D.Sonawane\textsuperscript{4}

\textsuperscript{1,2,3,4}Department of Information Technology Engineering, Matoshri College of Engineering and Research Centre, Eklahare, Nashik.

Abstract: Recently, Smartphone gains more and more popularity not because of its look and feel but peoples are addicted to use various applications of Smartphone. With the use of smartphone apps user also worried to reveal their identity while downloading and installing app in their mobile as many apps may require user information. For example, what’s app messenger, hike messenger etc. uses contact list information from user cell phone, Google map, drive apps like uber, Ola etc uses contacts, location information of user. Mobile apps mostly varied and less understood typically for their popular functionality. Users are neglecting to install apps as security and privacy point of view. To overcome security issues there is need of such system which gives recommended app list to the end user. User recommendation is based on popularity as well as security concern.

Keywords: Mobile Apps, Recommender Systems, Security and Privacy

1. Introduction

Generally, mobile apps are developed using computer programming languages such as, java, android, .net, X-code etc. Apps are designed in such way to match with device layout such as computer, mobile, tabs etc. After successful completion of application programs is using programming language, they have to plot into electronic devices like, mobile, tabs, android wear etc. Application distributed platform is used to relate device and develop apps. In digital market, there are lots of mobile apps are available, few of them are freeware and at the same time others are In-purchased apps. Usually, apps are downloaded and installed in target device such as, android, windows, iPhone etc. The app availability based on the public demand, so the designer tools herd fast and rapid extension of mobile app into various other kind such as games in mobile, factory computerization, services based on GPS and based on location, banking, ticket purchasing, etc. There is a challenging issue in mobile app recommendation, because of sudden increase in quantity and the variety of mobile apps which in turn led to the conception of broad range of review and creation sources including blogs, magazines and online app services [1]. Each and every day multiple app stores launched a daily app leader boards that outputs the app ranking chart as per app popularity. Leader board is one of the most popular ways to promote mobile apps [3]. Therefore, according to the ranking on leader board app get downloaded by mobile users. And this is the reason of more advertising of mobile apps by their developers. However, leafy App developers find some fraudulent means to purposely boost their Apps which eventually manipulate the chart rankings on an App store. Mobile app ranking is varied according to its functionality with respect to privacy and security [2].

To provide better experience for end user some apps used personal information of user such as, contact list, location based service to extract user’s current location. For instance, users may not expect to share their location as well as their other personal information. According to recent survey, it seems that multiple peoples were not interested to download the apps that required personal information of user and also many of them were uninstalling apps from their mobile phones which have functionality to access private information of user [4-6]. Therefore, there is a need of such a kind of system that provides recommendation of apps for end user with security and privacy perspective. We are going to propose a recommendation system which work against user privacy and secrecy and then recommend apps for end user. For implementation such kind of system we will have to evaluate risk score i.e. privacy factor. It may be security wise and popularity wise score [2].

However, for security and popularity of apps we need to explore app dataset and required to refine it. And have to evaluate certain kind of permissions as,
normal permissions having minimal risk to other applications. Dangerous permissions provides access for private data and finally signature/System permissions which required system signature certifications like, ability to control the overall process of the system. In app mining process we give preferences for rating based apps, review based app and also trying for survey based customer reviews. Which can further combine together for generate recommended app-list for end user. By evaluation of risk score and app mining method we will have to prove efficiency and capability of our system. Our final contribution is detection of ranking fraud detection and app recommendation.

2. Literature Survey

H. Zhu, et al [1], investigated three types of evidences such as, ranking based evidence, rating based evidence and review based evidences. In ranked based app detection there are three types of phases included such as, raising phase, maintaining phase and recession phase. App fraudulent is determined from ranking and maintained period. In Rating based evidence, comparative analysis is made between current app rating and historical app rating. Lastly, in review based evidence historical comments are analyzed and then app fraud is detected. TaintDroid is information flow tracking system introduced in [2]. There are 68-instances of potential misuse of user’s private information across 20 applications. This system monitors the actual nature of third-party android-applications. TaintDroid uses variable level tracking within virtual machine interpreter. In this system, multiple taint marking stored as taint tag. To minimize performance overheads TaintDroid system tracks explicit data and it does not control implicit flow. TaintDroid specifically designed for malware detection. It provides analysis of real-time application other third-party applications by monitoring the behaviour of apps. Basically, in android as well as iOS web view is main component.

D. Gleich et al [3], applied in recent algorithmic theory of matrix completion to skew symmetric matrices. The algorithm matrix completion handles the skew-symmetric data which is then used to extract ranks for each item. It can be applied in pairwise comparison and rating data as it is robust for noise and incomplete data. Existing Kemeny optimal ranking techniques are NP-hard. Therefore, in large scale rank aggregation settings it is an appropriate. In this proposed scheme measures the pairwise scores, solves the matrix completion problem which determines the quality of items. It only required to resolve convex optimization problem with a unique global minima. SVP solver is adapted for matrix completion. The solver evaluates an explicitly chosen rank.

L. Azzopardi et al [4], discussed about language model for information retrieval. Unigram based LM is latent variable which successfully applied to IR. It is known as, probabilistic latent semantic indexing (PLSI). M. N. Volkovs et al [5], formulate a flexible probabilistic model. They addressed the problem of where preferences are expressed in other forms. In this consistency exist in observed preferences. Pairwise comparisons are the building blocks of all general model evidence. They have addressed the problem constructing a flexible probabilistic model over pairwise comparisons. A score based approach is adopted for rapid learning and inference by which model get applicable in large-scale problems with hundreds of thousands of preferences. For future work they have predicted developing supervised extensions of the model which can be more directly utilize the labeled training data available in problems such as Meta search.

Y. Ge et al [6] introduced a parameter based scheme to mine evidences of travel route. A route mark is discussed to represent a typical driving path from one interesting site to another. A generative statistical model is characterized to the distribution of distance and identification of evidences of driving distances. In this they have evaluated taxi driving fraud detection system with large scale real-world taxi GPS logs. They encode trajectories using symbol and oversampling. Lastly, they have provided case study by exploiting real-world GPS logs of around 500 taxi drivers during the period of 30 days. With all this, some interesting regularities of taxi driving fraud activities are detected.

A. Klementiev, et al. [7], demonstrated an effectiveness of ULARA on data fusion task across ad hoc retrieval systems. ULARA is introduced to solve the problem of unsupervised learning algorithm. It is a generalized reranking approach. In synthetic data task, ULARA can not only distinguish good ranking functions. In inductive learning, many applications required ranking expressivity. The problem of rank aggregation is one in which multiple existing rankings of an item set are combined into a joint ranking. The data fusion problem is occurred in information retrieval community which corresponds to document ranking based on input of several IR systems. ULARA is linear combinational ranking functions which are guided by effective principle. It is the first method to learn parameterized rank aggregation without supervision.

B. Zhou et al [8] represented the notion of spam city. It is basically used to identify spam pages. It is the more flexible and user-controllable measure than the traditional supervised classification methods. The proposed method is cost effective in which real data set is used to evaluate an effectiveness of methods. With the experimental set up it is analyzed that the proposed method is effective for the detection of spam pages. A local greedy method is used to search
spam pages which extracts the whole (θ, k) farm of target page.

H. Zhu et al. [9], proposed an approach to enrich to the contextual information of apps [12]. All these contextual information is combined together into maximum information entropy model for training mobile app classifier. A leverage Web knowledge is taken into consideration to enrich to textual information of app. In this approach app name is submitted to the web search engine and then search snippet is obtained as additional textual information of corresponding app. A search snippet can be defined as abstract of web pages which gives the relevant result to the search query. Many classification approaches have been discussed in this paper such as, Naive Bayes, SVMs and Maximum Entropy (MaxEnt) etc. A leverage MaxEnt is proposed for mobile app classifier. To select effective feature functions MaxEnt model is an effective function.

H. Zhu, et al [10], discussed about an approach to enrich the contextual app information by discovering more and more web knowledge from web search engine. Classification of app consists of two phases such as, in first phase gathered many context logs from mobile users and then discovered features of web based knowledge and real world contextual features of apps.

A. Mukherjee et al [11], proposed an unsupervised model. In this model spams are formulated as an unsupervised problem. Also there evaluation is based on judgement of human experts. It is used in prior work of opinion spam. It enables analysis of posterior density and detection in individual framework. In this paper, a novel is provided to evaluate the results of unsupervised opinion spam models using supervised classification without the need of any manually labeled data.

3. Proposed System

1. Mining App Information:

At the beginning, in app mining phase system will discover leading event’s from the historical ranking records of apps. Then it will integrate adjacent leading events for construction of leading app sessions.

2. There are three types of evidences investigated in previous literature survey such as, ranking based evidences, rating based evidences and review based evidences. Each evidence is explained as below:

a. Ranking Based Evidences:

There are three different phases included in the analysis of ranking based evidence such as, rising phase, maintaining phase and recession phase.

- In rising phase, app ranking increases peak point position in leader-board specifically, in each leading event.
- In maintaining phase, this peak position is maintained for specific time period and
- In recession phase, it is decreases till the end of event.

Ranking range of app is depending upon the beginning and end time of event i.e. maintaining time. Leading session of apps has ranking fraud, if their behavior in these three ranking phases of leading events should be different from those in a normal leading session. System actually identifying each app with ranking manipulation always has an expected ranking target and the hired marketing firms also charge money according to such ranking expectation. The suspicious or fraudulent leading events contains very short period of rising and recession phase. In the intervening time, the cost of ranking manipulation with high ranking expectations is quite expensive due to the unclear ranking principles of App stores and the fierce competition between App developers. Therefore, fraudulent app could be the app’s having very short maintaining phase with high ranking positions.

b. Rating Based Evidences:

It is beneficial for ranking fraud detection. Sometimes, ranking based evidence is not sufficient


