DETECTING FRAUD APP USING SENTIMENT ANALYSIS

Salini[1], Dhevadharshini[2], Malathi[3]

Student, Anand Institute of Higher Technology, Chennai, India[1, 2]

Assistant Professor, Anand Institute of higher Technology, Chennai, India[3]

Abstract:Fake Web application   is software that mimic functionality of valid, reliable and genuine applications. once this applications are installed they perform malicious actions like aggressively display advertisements to get revenue from that, intercept sensitive data from your system, infect devices and so on. Most of the times the user cannot differentiate between the fake and legitimate applications hence before downloading any app people always enquire about the opinion of the app by the users. In this paper we are introducing a platform where people can enquire about the application before downloading it. The results are based on previous reviews and ratings given by the user which   provides a chance to recognize user experience with a particular mobile app. Basically we will analyze the reviews using sentiment analysis which is a text classification technique which analysis text and tells whether the sentiment is positive, neutral or negative

Keywords: Sentiment analysis, text classification

I. INTRODUCTION

.

Positioning misrepresentation for versatile application showcase alludes to fake or tricky exercises which have a reason for knocking up the applications in the prominence list. It turns out to be more continuous for application designers to utilize shady means, for example, expanding their applications deals, to submit positioning misrepresentation. We give all encompassing perspective of positioning misrepresentation and propose a positioning extortion identification framework for versatile applications Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. A sentiment analysis system for text analysis combines natural language processing (NLP) and machine learning techniques to assign weighted sentiment scores to the entities, topics, themes and categories within a sentence or phrase

II. ANALYSIS

OBJECTIVE

design a system this may detect fake apps by considering different evidence indicating their true behavior, To find apps are real or not, To increase the classification accuracy of a system.

The main objective of the proposed work was a review fraud detection of apps and to use sentiment analysis approach to differentiate the particulaTo r fraud apps

Through as we found that the proposed method gives accuracy of 90 % in comparison with other algorithms.

SCOPE

This project has more scopes as it would be more beneficial for the villagers as they do the manual works online. Following are the scope of the project.

•. Scope App Store Analysis writing   incorporates that performance investigation on a gathering of applications rates ponders   mined from an App Store.

SYSTEM ANALYSIS

. This is the process of collecting and interpreting facts,identifying the problems and decomposition of a system into its components.it is conducted for the purpose of studying a system or its parts in order to identify its objectives.

Problem Definition

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from

many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Existing system

In the literature, while there are some related work, such as web ranking spam detection, online review spam detection and mobile App recommendation, the problem of detecting ranking fraud for mobile Apps is still under-explored Generally speaking, the related works of this study can be grouped into three categories.The first category is about web ranking spam detection.The second category is focused on detecting online review spam.Finally, the third category includes the studies on mobile App recommendation

Proposed system

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. We further propose two types of fraud evidences based on Apps' rating and review history, which reflect some anomaly patterns from Apps' historical rating and review records.In Ranking Based Evidences, by analyzing the Apps' historical ranking records, we observe that Apps' ranking behaviors in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession.

SYSTEM DESIGN

IV.MODULES

.
- ➢ Mining Leading Sessions
- ➢ Ranking Based Evidences
- ➢ Rating Based Evidences
- ➢ Review Based Evidences
- ➢ Evidence Aggregation

Mining Leading Sessions

In the first module, we develop our system environment with the details of App like an app store. Intuitively, the leading sessions of a mobile App represent its periods of popularity, so the ranking manipulation will only take place in these leading sessions. Therefore, the problem of detecting ranking fraud is to detect fraudulent leading sessions. Along this line, the first task is how to mine the leading sessions of a mobile App from its historical ranking records. There are two main steps for mining leading sessions. First, we need to discover leading events from the App's historical ranking records. Second, we need to merge adjacent leading events for constructing leading sessions.

Ranking Based Evidences

In this module, we develop Ranking based Evidences system. By analyzing the Apps' historical ranking records, web serve that Apps' ranking behaviors in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase. Specifically, in each leading event, an App's ranking first increases to a peak position in the leaderboard (i.e., rising phase), then keeps such peak position for a period (i.e., maintaining phase), and finally decreases till the end of the event (i.e., recession phase).

Rating Based Evidences

In the third module, we enhance the system with Rating based evidences module. The ranking based evidences are useful for ranking fraud detection. However, sometimes, it is not sufficient to only use ranking based evidences. For example, some Apps created by the famous developers, such as Gameloft, may have some leading events with large values of u1 due to the developers' credibility and the "word-of-mouth" advertising effect. Moreover, some of the legal marketing services, such as "limited-time discount", may also result in significant ranking based evidences. To solve this issue, we also study how to extract fraud evidences from Apps' historical rating records.

Review Based Evidences

In this module we add the Review based Evidences module in our system. Besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Indeed, review manipulation is one of the most important perspective of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users often first read its historical reviews to ease their decision making, and a mobile App contains more positive reviews may attract more users to download. Therefore, imposters often post fake reviews in the leading sessions of a specific App in order to inflate the App downloads, and thus propel the App's ranking position in the leader board.

Evidence Aggregation

In this module we develop the Evidence Aggregation module to our system. After extracting three types of fraud evidences, the next challenge is how to combine them for ranking fraud detection. Indeed, there are many ranking and evidence aggregation methods in the literature, such as permutation based model score based models and Dempster-Shafer rules . However, some of these methods focus on learning a global ranking for all candidates. This is not proper for detecting ranking fraud for new Apps. Other methods are based on supervised learning techniques, which depend on the labeled training data and are hard to be exploited. Instead, we propose an unsupervised approach based on fraud similarity to combine these evidences.

V. RESULTS AND DISCUSSION

In this project, we developed a ranking fraud detection system for mobile Apps. Specifically, we first showed that ranking fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. Then, we identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. Moreover, we proposed an optimization based aggregation method to integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps. An unique perspective of this approach is that all the evidences can be modeled by statistical hypothesis tests

VI. CONCLUSION

Sentiment analysis it is easy to be extended with other evidences from domain knowledge to detect ranking fraud. Finally, we validate the proposed system with extensive experiments on real-world App data collected from the Apple's App store. Experimental results showed the effectiveness of the proposed approach. In the future, we plan to study more effective fraud evidences and analyze the latent relationship among rating, review and rankings. Moreover, we will extend our ranking fraud detection approach with other mobile App related services, such as mobile Apps recommendation, for enhancing user experience.

REFERENCES

[1]Hengshu Zhu, Hui Xiong, Senior Member, IEEE, Yong Ge, and Enhong Chen, Senior Member, IEEE ―Discovery of Ranking Fraud for Mobile Apps‖ IEEE Transactions On Knowledge And Data Engineering, Vol. 27, No. 1, January 2015

[2] Pranjali Deshmukh, Pankaj Agarkar ―Mobile Application For Malware Detection‖ International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 02 | May-2015 www.irjet.net

[3] Anuja A. Kadam ,Pushpanjali M. Chouragade ―A Review Paper on: Malicious Application Detection in Android System‖ International Journal of Computer Applications (0975 – 8887) National Conference on Recent Trends in Computer Science & Engineering (MEDHA 2015