DISCOVERY OF FRAUD APPLICATIONS IN MOBILE STORES

Shruthi[1] and Prof.Anand s uppar[2]

[1]Department of Computer Science

[2]HOD,Department of Computer Science

ShreeDevi Institute of Technology

Kenjar,Mangalore,575015,

India

ABSTRACT

As users of mobile phones and mobile stores, most of us have become the victims of fraud or deceptive activities by downloading or purchasing some applications which is of no use from mobile stores after viewing some fake ranking and also after reading some fake reviews. As we all know that numbers of mobile applications are increasing day by day and the question arising here is nothing but as the count of these mobileapplications increases why the number Frauds are also increase. The answer is simple that as new applications comes to market either for the purpose of stepping up or bumping up the applications or for the purpose of hitting down the applications from the popularity list in the mobile stores the third parties or the hackers and even the developers do some fraud activities by providing some fake ranking and even writing some fake reviews. The app developers or the frauds use many shady methods including fake reviews and rating aiming at the promoting or demoting applications from popularity list in mobile stores and in my research i am going to show how to find whether an application in mobile store is an fraud or genuine one by having look on to large pre-existing data base or in other word we can say that by analyzing parameters or evidences from data bases such as rank based evidence, review based evidences, rating and even analyzing some malicious links.

Key words: App data , *Fraud detection, Fake ranking.*

1. INTRODUCTION

Discovery of Fraud application in mobile Store which is done using data mining as a base will help you to discover fraud applicationsand genuine applications in the mobile store by examining all the evidence or parameters stored in the data base such as ranking, rating and also malicious links based evidences. Many researches have been done in this field in order to discover the fraud applications in the mobile store and it is still under explored or renowned because of many reasons. Here we are trying to improve parameters for discovering of fraud applications by using ranking, review and also malicious link based evidences. Here we are also looking on to statistical or graphical view of ranking and reviews based evidences provided by all its users including frauds and there by detecting whether an application in the mobile store is fraud application or not.

2. BACKGROUND AND RELATED WORKS

First survey is based on the paper [1] finds all the ranking frauds of applications in the play store of your mobile. It uses the concept of local and global analomy where local analomy is the fraud for giving fake ranking to the applications in the mobile and global analomy may be an uploader or the developer of the applications. Global analogy can upload the applications after he had logged in using his/hers username and password and it is also possible for the person to view all the user details. The researcher of this paper claims that user can not only view the global and local ratings but also it is possible for a user to view download ratings. They claim that the user can view all the applications uploaded and can download the applications he required and also possible to rank the applications based on his satisfaction after using the application.

Next paper i have chosen for contemplate is [2]. Let us consider some of the example of areas of studies such as information retrieval, collaborative studies and social choices. All these examples faces a common problem which is nothing but preference aggregation in which there will be multiple preference for among these object which should be combined to perform solidarity ranking. Each of these preferences can be expressed in variety of forms but it makes the aggregation more complex. In order to solve these problems we are introducing a model called Flexible Generative Model which solves all the problems by making pair wise comparisons. The main advantage using this model is very fast inference which results in applying preference aggregations to any number of cases.

[3] is next paper i have chosen for evaluation. Here in this article they have made analysis on some latent or hidden factor known as pureSVD, a memory based model and on a PCA based model using evaluation criteria as variety and accuracy. The analysis result is nothing but due to its trust on explicit feedback the latent factor or pureSVD did not perform well and in case of memory based model what happen is that only for popular applications in the high dimensional space, they perform vector operations as they could not make grade to capture neighbors of less popular applications.

The main pros of this is that they provide high accuracy due to the vector operations they perform but in case of variety they perform poorly due to the failure of capturing neighbors of less popular applications.
One important paper we are going to inspect is nothing other than [4]. First let us have a look on to what is web spam. For the purpose of influencing results of search engines like Google what have been done is that artificially created pages are inserted to web for the purpose of driving traffic to some pages for fun or for some profit which results in decreasing quality of the results for what we searched. And now let us have a look on to kinds or types of web spam and their detecting methods.

The first spam or in simple terms we can call it as superfluous or unwanted mass messages we are going to discuss is nothing but the one that is most commonly found email that is" email spam". Email spam is an electronic spam where irrelevant messages are sent especially done for some sort of advertisements where the carrier of this message is email. The next type of unwanted message is content based web spam is a malicious activity by which content of a page is infected for the purpose of obtaining higher rank. Content spamming methods are of different type such as body spam, tittle spam, Meta tag spam, URL spam and so on. First one is Body spam is a type of spamming method in which the spam is placed somewhere in the body of the document. Next title spam is a type of content spamming method in which the spammer fill title of document or the report with unconnected words because search

International Journal of Advances in Scientific Research and Engineering (ijasre)
ISSN: 2454-8006
[Vol. 03, Issue 4, May -2017]
www.ijasre.net

engine sometimes gives higher weight for this tag. Meta tag spam is another spamming method in which unrelated words are placed in the HTML Meta tag.

There are various methods for detecting spam including machine learning method, link based method and so on and the method used in this survey for detecting web spam is content analysis in which it instinctively notices all the spam pages by some formerly unspecified techniques and also checks the efficiency of these methods**.**

Next appraisal we going to consider is [5]. Many individuals and business organization are now a day depends on opinionated social reviews for making decisions about various products for their business promotions. For example consider online shopping site in which if the product is of no use or if from the customer side if the company got negative feedback of the product then the company will no longer allow it to be available in their sites as selling of bad quality products may effect business as well as it may reduce the reliability of customer on that site. As in case of mobile applications what people do is for the purpose of bumping up or demoting some product they do opinion spamming for example by fake reviewing ranking etc. In order for finding such spams many methods have been suggested and here a model called Author spamicity model have been suggested.

The working of the model is based on Bayesian networks and it will make use of different observed behavioral foot print of reviewer like malicious links in the reviews etc.

The last paper that we are going over is[6]. Let us take an example of some sort of object in which a group of judges where frequently asked to rank these objects based on some criteria or some benchmarks and here we assume that their judging skills depends on some object's domains. While comparing with other technologies the main problem faced byInformation retrieval and natural language processing is learning to rank a group of objects with a purpose of producing better joint ranking. Here they introduced some framework for the purpose of grouping of the entire ranked object with some object's domain based judgment without supervision or surveillance as if supervised ranking data coming from multiple domain islittle hard to obtain. Here what they have done is that they applied learning framework to the set of all ranking object and then aggregated top k list of objects showing all the critical improvement over a domain agnostic baseline for both cases including supervised and un supervised ranking cases.

3.EXTRACTING EVIDENCE FOR RANKING FRAUD DETECTION

3.1Mining leading session

We know data mining is inspecting or examining pre-existing data base so as to generate some new information's that is analyzing data base to extract some information's which can be used to solve problems by transforming the extracted information in to some understandable structures or forms . Here Mining Leading session is nothing but sessions which is to be taken for mining or for analyzing. As here we are dealing with detections of frauds in mobile applications the leading session will be some applications in the App stores.

Leading sessions depicts period of popularity of a mobile applications and hence this session is responsible for handling of all evidences such as ranking, rating and so on. Thus by having diagnosis on this sessions we can detect the frauds in a mobile applications in play store. First step in finding fraud is to mine the leading sessions from the

historical records of the mobile app. Mining leading session from historical record can be done in two steps .By finding leading events from the historical records of the mobile applications in the play store and secondly by merging these events for constructing mining leading sessions.

3.2 Ranking Based Evidence

In order to study ranking parameter we need to analyze the available historical records in which the ranking performance satisfies some specific ranking pattern. These ranking pattern patterns include three different phases especially rinsing phase, maintaining phase and recession phase and now let us have a look on to these phases. Increasing of applications ranking to crest positions in the leader board when a leading or an important event take place is called rinsing phase. As we discussed that if a leading event takes place the ranking of the application initially increases to peak or crest position and then it keep such peak position like that for a period of time then such a phase where the peak position is maintained for a short period of time is called maintaining phase. After short period of time peak positions decrease spontaneously till the event ends this phase is called recession phase.

3.3 Rating Based Evidence

Next Schedule or module we are going to discuss is Rating Based Evidence. We know that ranking evidence help us to find ranking fraud but in most of the cases these evidence is not enough for classifying whether a mobile application is a fraud or genuine one. Consider an example, one of the famous video game developer in Paris called Game loft. Sometimes applications produced by them may have some sort of leading events which includes large value of 0 and 1. This may happen because of developer's trustworthiness and may also happen due to word-of-mouth advertising effect. Another example for ranking based evidence is legal marketing services. To solve all these problems we need to study the fraud or deceptive activities from the historical rating records.

3.4 Review Based Evidence

Most of the applications allow textual comments as reviews for the users to write about what they feel after using particular applications or a product. Also most of the business organizations are now using this information to make decision such as whether to promote or demote some products. For example if the review of customers about a product after purchasing it from some shopping site is very bad ,they can remove that product from there sites because selling of bad product effect their business. Analyzing reviews are one of the most important evidence in fraud detection as the user usually reads the historical reviews for making decisions before purchasing or downloading particular mobile applications from the app store of particular mobile. People often prefer to buy or download mobile applications from the App sore of mobile applications having more positive reviews. Developers or frauds usually write fake reviews for attracting the users to download an application and even they may write negative reviews about a mobile application for the purpose of demoting it from its popularity list.

3.5 Evidence Aggregation

We have extracted three types of evidences including ranking, rating and review based evidences and we need to combine them in an effective way so as to find the frauds in mobile applications. There are many methods to aggregate all these parameters or evidences including permutation based models, score based models, dempster-shafer rule and many more. But one of the disadvantages of using these aggregation methods is that thy concentrate only on

learning of global ranking of the candidate and so these cannot be considered as good methods for finding fraud applications in mobile stores. There are some other techniques that depend on labeled training data but are hard to exploit which is based on supervised learning techniques. Hence we introduce a new technique for combing these evidences in an effective way based on unsupervised approaches.
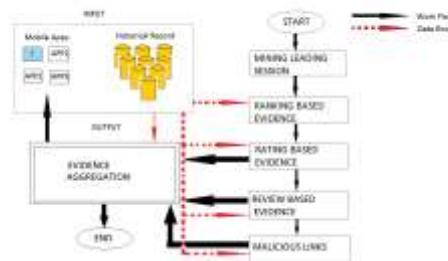
Fig.1.The framework of Discovery of Fraud Application in Mobile

4. DISCUSSION

Detecting frauds application's in mobile store is possible only by analyzing all available evidences such as ranking, rating, review evidences and then aggregating all the evidences to find frauds. But even though there are many evidences there is a chance of failure of detection of these kinds of frauds. By increasing or even by improving these parameters we can increase the chances of detecting frauds that is the probability of finding fraud is high in proportion to the rate of increase of detection parameters. Here for finding fraud applications in mobile store first we are looking on to rank based evidences if the rating of app is in such a way that if the rating is high when the app is uploaded and suddenly the rating goes down then we can say that it is a fraud app else if the rating is such a way that rating is first high when the app is uploaded then after that without suddenly going down if it maintain a minimum rating we can say that the application in the mobile store is a genuine one. Here I have also made some improvements to rating based evidence that is I have provided rating evidence with an additional statistical view so that by looking in to statistical view or graphical representations we can say whether the application is a fraud or genuine one. If the statistic or the graphical representation of the rating based evidence is in such a way that it first increases to peak or crest and suddenly falls down to the bottom then we can say that it is a fraud application or else the graph first increases to peak and then maintain a rating for period till the end then we can say that it is a genuine applications. There is one more case that is if the graph first increases to peak then maintains a rating for a period of time then suddenly decreases .In this case also we can say that it is a genuine applications. Next we can have a look on to what improvements have been made to review based evidence. Review based evidence is one of the most important aspect where most of the user make decision on whether to download or purchase the applications from the mobile stores or not. It is also used for promoting and demoting of applications from the mobile stores that by giving good reviews by the frauds or by the uploader they meant for the promotion of applications in the mobile store and if the frauds give all negative comments as reviews then it is meant for demoting the applications from its popularity list. Here also we are providing statistical view. Usually by reading all the reviews the customers or the users of mobile applications usually make decision of whether to download or purchase the applications from mobile store. If all the reviews say that

theparticular application is good the users of mobile stores download the applications from the mobile stores else there is a less chance for users to download particular mobile applications. In review based analysis there is a chance of providing some malicious link as reviews. Malicious link can be a link for downloading some applications. It can be some viruses also and thus by analyzing all these reviews we can detect frauds.

5. CONCLUSION

In this paper, 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, thus 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.

REFERENCES

[1]Hengshu Zhu, HuiXiong, Senior Member, IEEE, Yong Ge, and Enhong Chen, Senior Member, IEEE,"Discovery of Ranking Fraud for Mobile Apps", vol. 27, no. 1, january 2015.

[2] M. N. Volkovs and R. S. Zemel, "A flexible generative model for preference aggregation," in Proc. 21st Int. Conf. World Wide Web, 2012, pp. 479–488.

[3] 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.

[4] 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.

[5] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 632–640.

[6] A. Klementiev, D. Roth, K. Small, and I. Titov, "Unsupervised rank aggregation with domain-specific expertise," in Proc. 21st Int. Joint Conf. Artif. Intell., 2009, pp. 1101–1106.