International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) e-ISSN: 2395-5325
Volume2, Issue 3 (March '2016)

A HOLISTIC VIEW OF RANKING FRAUD DETECTION SYSTEM FOR MOBILE APPS

Mrs. N.Bhavani,
Department of Information technology
Saranathan College of Engineering,
Trichy-620012, Tamil Nadu, India.

A.Gayathri,
Department of Information technology
Saranathan College of Engineering,
Trichy-620012, Tamil Nadu, India.

T.Eswari,
Department of Information technology
Saranathan College of Engineering,
Trichy-620012, Tamil Nadu, India.

K.Kalaivani,
Department of Information technology
Saranathan College of Engineering,
Trichy-620012, Tamil Nadu, India.

N.Kamachi Swatika
Department of Information technology
Saranathan College of Engineering,
Trichy-620012, Tamil Nadu, India.

Abstract: Ranking fraud in the mobile App market refers to fraudulent activities which have a purpose of raise the apps in the popularity list. In fact, it is easy for App developers to boost their Apps sales or posting forged App ratings, to commit ranking fraud. To prevent this ranking fraud we provide a holistic view of ranking fraud detection system for mobile Apps. Specifically, we first propose to locate the ranking fraud by mining the leading sessions, of mobile Apps. Such leading sessions can be leveraged for detecting the abnormality or irregularity 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 addition, we propose an optimization based aggregation method to integrate all the evidences for fraud detection.

Keywords: ***Mobile Apps, ranking fraud detection, evidence aggregation, rating and review.***

I.INTRODUCTION

Most of us use android Mobile these days and also uses the play store capability normally. Play store provide great number of application but unluckily few of those applications are fraud. Such applications dose damage to phone and also may be data thefts. Hence such applications must be marked, so that they will be identifiable for play store users. To stimulate the development of mobile Apps, many App stores launched daily App leader boards, which demonstrate the chart rankings of most popular Apps. Indeed, the App leader board is one of the most important ways for promoting mobile Apps. A higher rank on the leader board 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 leader boards.

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

II.RELATED LITERATURES

A. FORENSIC DETECTION USING CLOUD STACK AND DATA MINING

This paper propose a system to develop a android application that will take reviews from two different websites for single product, and analyse them with NLP (Natural Language Processing) for positive and negative rating.NLP is the interaction between human and computer and it enabling man machine interaction through natural language . Cloud computing is the use of hardware and software computing property via internet. Load balancing technique is the process of improving presentation in cloud computing. It is used to control the

traffic across the connection. K-mean's algorithm is one of the data mining concept used for identifying the positive and negative reviews between two different websites. The process of collecting, searching and analysing a large amount of data in a database is called data mining. R provides a wide multiplicity of arithmetical and graphical technique. R language is used to generate the analysis graphs, like histogram, bar graph, pie chart. Using these techniques they detect the fraud apps among other apps over the internet.

B.HEALTH CARE DETECTION:

This paper introduces some preliminary knowledge of U.S. health care system and its fraudulent behaviours, analyses the characteristics of health care data, reviews and compares currently fraud detection approaches using health care data in the literature as well as their corresponding data preprocess methods. Antifraud approach incorporating data mining, machine learning or other methods known as Multilayer perception (MLP) neural network is a widely used because it has many advantages such as it can handle complex data structure especially non-linear relationship and it has high tolerance to noisy data. Another technique in health care fraud detection is decision tree.

Five approaches are used to construct the classifier: minimizing the false positive (FP), minimizing FP with misclassification weights, minimizing the false negative (FN), minimizing the false positive with balanced class in conjunction with misclassification weights and combining diverse classifiers together and the clustering model considering geo-location information of both Medicare/Medicaid beneficiaries and providers to flag suspicious claims. Data mapping map the Geo-graphic location to Medicare data. Calculate distance between beneficiaries and service providers then select the three common diagnoses after completing analysis for payment.

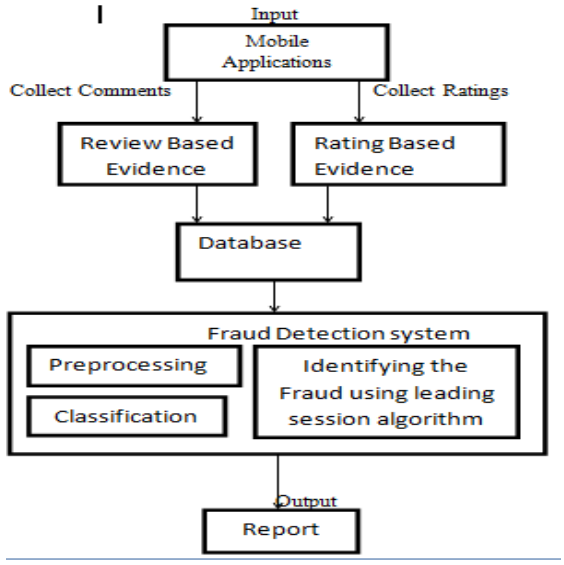
C. AUTO INSURANCE FRAUD DETECTION BY DATA MINING

The paper presents fraud detection method to predict and analyse fraud patterns from data. It evaluate data mining techniques to solve fraud detection in automobile insurance. There are a number of data mining techniques like clustering, neural networks, regression, multiple predictive models. Techniques used here are Bayesian network and Decision tree. Bayesian network is the technique used for classification task. Classification determines which of these classes a specific data belongs to. For the purpose of fraud detection, we construct two Bayesian networks to describe the behaviour of auto insurance. Bayesian network is constructed under assumption that the driver is fraudulent and other under the assumption that the driver is a legal. Descriptive models in Decision trees are created to describe the characteristics of fault. Decision trees express independent attributes and a dependent attribute in a tree-shaped structure. C4.5 Decision tree can help to make predictions and to extract crime patterns. It generates rules from trees and handles numeric attributes, missing values, pruning, and estimating error rates. However, performance decrease can occur when C4.5 is applied to large datasets.

D. REVIEW OF FINANCIAL ACCOUNTING FRAUD DETECTION BASED ON DATAMINING

Financial accounting fraud detection (FAFD) has become an emerging topic of great importance for academic, research and industries. To detect financial accounting fraud, collective known as forensic accounting Common classification techniques such as neural networks, K-nearest neighbour, and Fuzzy logic and Genetic Algorithm are used. Neural Networks-The neural networks can be used to predict the occurrence of corporate fraud at the management level .Nearest Neighbour Method- Is a similarity based classification approach. Based on a combination of the classes of the most similar k record(s), every record is classified. Sometimes this method is also known as the k-nearest neighbour technique. It is used in automobile insurance claims fraud detection and for identifying defaults of credit card clients. Fuzzy logic and Genetic Algorithm – Genetic algorithm along with binary support vector system (BSVS) which is based on the support vectors in support vector machines (SVM) are used to solve problems of credit card fraud that had not been well identified . Fuzzy Logic is a mathematical technique that classifies subjective reasoning and assigns data to a particular group, or cluster, based on the degree of possibility the data has of being in that group. This paper suggests that using only financial statements data may not be sufficient for detections of fraud.

IV. SYSTEM ARCHITECTURE



APP SELECTION PROCESS

In App selection module we choose the app process role, if the entry user has want to do the app view , giving rating or review otherwise select best app for their personal use. If the entry has come to do the admin process they should give their authentication for further proceed, but this authentication is not checked in user gateway entry. The system process has follow, which person come to do the activity in our project and what they done, in the mining session keep and record all those information, if the action has commit the valid information in this process.

RATING BASED EVIDENCE

The Rating based evidence model has useful for ranking fraud detection. We also study how to extract fraud evidences from user rating records for Apps. Specifically, after an App has been published, it can be rated by any user who downloaded it. more users to download and can also be ranked higher in the leader board. Intuitively, if an App has ranking fraud in a leading session, the ratings during the time period may have anomaly patterns compared with its user ratings, which can be used for constructing rating based evidences. In our project verify the user id for which is come to give the rating for particular app.

REVIEW BASED EVIDENCE

Most of the App stores also allow users to write some textual comments as App reviews. Review manipulation is one of the most important perspectives of App ranking fraud 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 download, and thus propel the App's ranking position in the leader board. In this module, the user gives review for particular app and it is stored in database after that, threshold value is calculated based on user's review. In this, post tagging, sentiment word net tools are used to filter the sentences for taking a sentiment word such as good, nice, bad, worst etc.

ADMIN ANALYZE RANK EVIDENCE AGGREGATION

The Apps ranking behaviours in a leading event always satisfy a specific ranking pattern. In this phase the admin has taken the rating and review aggregation for find the fraudulent app. Finally, admin draw an graph for rating and review. There are many ranking and evidence aggregation methods in the literature such as permutation based models score based models and Dumpster Shafer rules.

SUMMARY

To prevent the ranking fraud we provide a holistic view of ranking fraud detection system for mobile Apps. Specifically, we first propose to locate the ranking fraud by mining the leading sessions, of mobile Apps. 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 test. Moreover, we proposed an optimization based aggregation method to integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps.

CONCLUSION AND FUTUREWORK

In this paper, we developed a ranking fraud detection system for mobile Apps. We propose a methodology to evaluate the security of Android mobile apps based on data mining. Then, we identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. An unique perspective of this approach is that all the evidences can be modelled 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.

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, HuiXiong, 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] PatilRohini, Kale Pallavi, JathadePournima, KudaleKucheta, Prof.PankajAgarkar."MobSafe: Forensic Analysis For Android Applications And Detection Of Fraud Apps Using Cloud Stack And Data Mining "

International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 10, October 2015 .

[3] Qi Liu, MiklosVasarhelyi "Healthcare fraud detection: A survey and a clustering model incorporating Geo-location information" 29th World Continuous Auditing and Reporting Symposium(29WCARS), November 21-22, 2013.

[4] Rajashree S. Jadhav , Prof. Deipali V. Gore, The author has proposed "A New Approach for Identifying Manipulated Online Reviews using Decision Tree". (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 1447-1450.

[5] RekhaBhowmik"Detecting Auto Insurance Fraud by Data Mining Techniques".Journal of Emerging Trends in Computing and Information Sciences. Volume 2 No.4, APRIL 2011 ISSN 2079-8407.

[6] Anuj Sharma -Information Systems Area Indian Institute of Management, Prabin Kumar Panigrahi - Information Systems Area Indian Institute of Management these authors have proposed "A Review of Financial Accounting Fraud Detection based on Data Mining Techniques".

[7] (2014). [Online]. Available: http://en.wikipedia.org/wiki/cohen's_kappa.

[8] (2014). [Online]. Available: http://en.wikipedia.org/wiki/information_retrieval.

[9] (2012). [Online]. Available: https://developer.apple.com/news/index.php?id=02062012a.

[10] (2012). [Online]. Available: http://venturebeat.com/2012/07/03/apples-crackdown-on-app-ranking-manipulation.

