AN OVERVIEW ON IDENTIFICATION OF FRADULENT AND GENUINE USERS IN ONLINE ADVERTISING MARKET

Karthik.M¹, Sasi Kumar.E², Sathish Kumar.R³ and Mani.A⁴
¹, ², ³ U.G.Student, Computer Science and Engineering, S.A.engineering College
⁴ Professor, Computer Science and Engineering, S.A.engineering College

Abstract—Deceitful or fly-by-night activities are increasing with untrusted data in the application that are referred to as fake data in mobile application. In fact, it is very common for app developers to make use of shaded resources, such as inflating them to commit fraud. Many precautionary measures have been proposed but still, there is limited perceptive and exploration in this area. Deceitful activities are ranked by mining the active periods, namely according to the post (data), of mobile Application. Furthermore, an approach can be designed to find genuineness of user using data mining in the cloud database and indicate to the user according to the results. Thus this system will be more effective and scalable for the detection algorithm in addition to some regularity of swindle tricks.

Key words — Cdear, SentReP.

I. INTRODUCTION

To buy and sell the used products the user need a proper platform where they can sell their valuable products with no intruders and fraud activities. The platform will be user friendly for the client where they can easily access the platform in anywhere any time with available user-friendly gadget (smart phone). Internet based applications such as Quickr, suleka, Olx, flipkart, Amazon, PayPal and eBay have become an important means for fulfilling social, shopping, advertising and money-transfer needs. Suspicious users such as deceivers, spammers and other types of attackers are progressively engaging in fraudulent activities. So this process to find those intruders in the online advertising market and publish the advertisement in ease. The buyer and the seller will be genuine to advertise their respective advertisement. They can also ping and call the seller or buyer in one to one conversation. The advertiser can also upload their images of the products and can give detailed description in quotes to create the impression towards the buyer.

In existing system, the Online Commerce System, a dealer or broker can post data as individual and there is no mechanism to detect it. The disadvantage of his system is can’t find Genuine Users, trust ability of application will be affected; Possibility is high for fake data.

Android is an OS created by Google designed for touch screen devices, mobile devices, such as smart phones and tablets. Android OS, you can customize your device in many ways. The Os is based on the Linux kernel and it is developed with the IDE such as eclipse, Android studio etc. Android Os is the largest installed base. The android application is developed with the help of some languages like xml for the frontend (GUI) and java or other languages for backend. The application can be published in the platform of Play store, apps9 etc.

Hence a system can be proposed with an approach to find a proper user, by using data related operation into cloud database and indicate to the user according to the results. These experimental results will be more effective and scalable for the detection of algorithm in addition to some regularity of swindle tricks.
II. RELATED WORK

In this paper [1], the dataset based on fraudulent activities using an algorithm and performance measured from almost all published technical and review articles in automated counting fraud detection. Most researchers regarded fraud users and non-fraud users or intruders as data subjects; eigenvalue covers auditor data, company governance data, financial statement data, industries, trading data and other categories. Almost all data in earlier research were of auditing data and later research established model by using sharing data and public statement data. The ratio data is more effective than accounting data, using mining algorithms including statistical test, regression analysis, neural networks, decision tree, Bayesian network, and stack variables etc. Regression analysis is used for hiding data. Thus the author concluded that this model is better for detecting than auditing based detecting rate without assisting and also mentioned the need to introduce other algorithms for no-tag data mining.

In this paper [2], Zhiguo Wang briefs about the existing sub graph extraction techniques are based on a complete clustering of the graph nodes, and the author’s proposed algorithm takes into account the fact that not every participating node in the network needs to belong to community. Also, presents a technique to find a set of dense sub graphs in a given sparse graph. The main application of this “dense sub graph problem,” is they are interpreted as communities, e.g., social networks. The problem for finding dense sub graphs helps to analyze graph structures and complex networks, this process is challenging. This problem is very similar to the problem of reordering matrices in sparse matrix techniques. The author exploits these similarities and adopts new idea, that is recognizing matrix column similarities, concerning to compute a partial clustering of the vertices in a graph, where each cluster represents a dense sub graph. The main advantage is that this method does not require specifications of the number of clusters as it is difficult to estimate in advance. The effectiveness of this method is demonstrated in a few real-life examples and computational process is also very efficient.

In this paper [3], Meng Jiang et al. have discussed about importance of web applications such as Amazon, Twitter, Face book, and Hotmail as it is important means for shopping tasks, social, information seeking, and working. Many incredulous users such as fraudsters, spammers and other types of attackers are progressively trying to engage in crooked activities such as faking popularity in political campaigns and scamming money out of Internet users. Favorably, many commercially available suspicious behavior-detection techniques are eliminating a large percentage of fraud, Sybil and spam attacks on very popular staging. The stakeholders only want to ensure that any behavior happening on them involves a real person interested in interacting with a specific Face book page, following a specific Twitter account, or rating a specific Amazon product.

In this paper [4], Buddhika H. Kasthurirachchy et al. have used detection and sentiment classification to brief about domain adaptation. Also discussed about the usage of social networks and micro-blogs, which is quite limited without sophisticated framework for mining and analyzing user’s opinion. “What other people think” can be predicted with these. The human produced baselines are inferior to supervised classification methods; hence using it for a framework is impractical. It is mainly due to the fact that the data have varying spans as it is from heterogeneous sources. Hence, the main feature for every useful framework is domain adaptation. The existing works on domain adaptation are also mentioned by the authors in this paper. Set of experiments were performed by them based on important parameters to check accuracy level of domain adaptation.

In this paper [5], the Neil Shah et al. have proposed FBOX, an algorithm intended to stealth attacks, catch small-scale that slip below the radar. The algorithm has the some desirable properties like theoretical underpinnings and it is shown to be highly effective on real data. It is more scalable and also evaluates FBOX on a large, public edge like who-follows-whom, social graphs from social networks with high precision to find many suspicious accounts which are persisting without any deferment. The
method to detect suspicious users in large online networks is also explained by the authors. Spectral techniques are being used by typical methods for catching these suspicious users.

In this paper [6], Koji Maruhashi et al. proposed a useful tool that visualizes and detects a new sub graph patterns within a local group of nodes in a heterogeneous network. The tool is called MultiAspectForensics. Existing approaches which are successful for small homogeneous network data are likely to fall short. They proposed a method with three data sets taken from different application domains and discussed the insights derived from these patterns. The algorithm based on scalable tensor analysis procedures, captures spectral properties of network data and reveals informative signals for subsequent domain-specific study and investigation, such as suspicious port-scanning activities in the scenario of cyber security monitoring.

In this paper [7], Tanya Gera et al. have presented a methodology that uses metadata about reviews for distinguishing suspicions. Metadata about reviews is private data like MAC address, IP Address, location, date-time and browser id. In recent years email spam and web spam are the two essential highlighted social issues. Sellers are misusing customer’s review and interest in online shopping. The review section transformed into a true center for review spammers to betray customers by making false studies for target things. Many techniques have been given by specialists to accord with this problem. They said that recognition of reviews spam using networking parameters and geographical statistics has not yet been reported.

In this paper [8], Muyuan Wang et al. have proposed a new routing scheme that makes use of the social structure of underlying 'owners' and their mobile devices in intermittently connected Networks. They designed this routing scheme by utilizing the social contact information. For example, the contact between device and owners exists only when they are attending regular meetings. The routing optimization problem was formulated by them, by giving priority to the ever-changing network connectivity. Moreover they have used an algorithm to extract contact information from social structure. They developed a routing protocol for intermittently connected networks and their performance is tested with real network traces and able to achieve a good delivery rate and latency. They concluded with comparison study and proved their results outperformed first contact protocol.

In this paper [9], Snehasish Banerjee et al. have discussed about strategies that are used for writing fake reviews. As a part of research the authors made an experiment by inviting participants to write fake reviews for hotels and analyzed their strategies. After which, they had a questionnaire session with the participants about their strategies. Experimental results showed that writing of fake reviews starts with the information gathered through common search engines, e.g. Google and review websites, e.g. Trip advisor. With this information short, catchy and concise fake review titles are provided.

In this paper [10], Robert Layton et al. make use of information obtained from online review sites where people share their information on products or services. The online review sites are noticeable for their reliability, but have been criticized for not ensuring the integrity of the reviews. A major problem with these sites is that people who write reviews (review hoax) are paid by owners of the product to write favorable reviews. In this paper, the authors proposed a method to find whether a review is genuine or not. They proved this as a useful method with their experimental results.

In this paper [11], Aakas Zhiyuli et al. have designed a novel system to detect fake reviews called Cdear. It is based on sentiment analysis on attributes of products. Primarily, they tried to capture the sentiment diverse of attributes of products from different consumers. They have conducted various experiments on the real life datasets and have employed 20 experts to assess the reliability of reviews by carrying out a simulation of shopping online to evaluate their effectiveness of their system. Also, they developed an evaluation system to help those experts to assess the reviews to ensure the accuracy of experiments. Finally they made a comparison study between the automatic results of their proposed system and human evaluated results, which proved that their system’s review matches approximately with human perception of reviews’ and reliability is 82.66% accuracy.
In this paper [12], Hamzah Al Najada et al. have used supervised classification methods to differentiate spam reviews from non-spam reviews. As a part of their research, they used trained classifier to detect spams. While training their classifiers they faced dataset imbalance issue because a spam review was only of a small portion of total review report. Hence this leads to poor performance, but has high level of accuracy. Later they overcame this problem by using bagging based approach to provide balanced datasets for training and detecting spam reviews. Their experimental results have outperformed the baseline methods that are used for spam detection.

In this paper [13], Veena H Bhat et al. have proposed a model SentReP. It is classification of Movie Reviews with Sentiment using Efficient Repetitive Pre-processing. It is a focused pre-processing technique to classify opinions and based on tested parameters. They worked on data set obtained from the Cornell Movie review. In this work the authors have proved the effectiveness and accuracy of SentReP with various volumes of data set, when compared to other existing approaches. Hence they concluded that this approach is very efficient in analyzing sentiments of movie reviews.

REFERENCES

[1] Shiguo Wang, “A Comprehensive Survey of Data Mining-based Accounting-Fraud Detection Research"IEEE, July-2010.
[2] Zhiguo Wang. Dense Sub graph Extraction with Application to Community Detection” IEEE, February-2012.
[3] Meng Jiang, Peng Cui, Christos Faloutsos,"Suspicious Behavior Detection: Current Trends and Future Directions”, Volume: 31, Issue-1, Jan.-Feb. 2016.
[4] Buddhika H. Kasthuriarachchy, Kasun De Zoysa and H.L. Premaratne,”A Review of Domain Adaptation for Opinion Detection and Sentiment Classification”, IEEE, June-2012.
[5] Neil Shah, Alex Beutel, Brian Gallagher, Christos Faloutsos,"Spotting Suspicious Link Behavior with iBox: AnAdversarial Perspective", ICDM, October-2014.
[6] Koji Maruhashi, Fan Guo, Christos Faloutsos,"MultiAspectForensics: Pattern Mining on Large-scale Heterogeneous Networks with Tensor Analysis", IEEE, August-2011.
[7] Tanya Gera, Deepak Thakur, Jaiteg Singh,"BILD Testing for Spotting Out Suspicious Reviews, Suspicious Reviewers and Group Spammers",IEEE,October-2015.
[8] Muyuan Wang and Klara Nahrstedt,"Social Structure Based Routing Of Intermittently Connected Network Using Contact Information", IEEE, January 2009.
[9] Snehasish Banerjee, Alton Y. K. Chua,"Understanding the Process of Writing Fake Online Reviews", IEEE, December 2014.
[10] Robert Layton, Paul Watters, Oana Ureche,"Identifying Faked Hotel Reviews Using Authorship Analysis”, IEEE, March 2014.
[11] Aakas Zhiyuli, Xun Liang, Yige Wang,”Discerning the Trend: Concealing Deceptive Reviews”, IEEE, January 2016.
[12] Hamzah Al Najada and Xingquan Zhu,"iSRD: Spam Review Detection with Imbalanced Data Distributions”. IEEE, March 2015.
[13] Asha S Maneck Pallavi R P, Veena H Bhat, P Deepa Shenoy, M. Chandra Mohan, Venugopal K R, L M Patnaik, "SentReP: Sentiment Classification of Movie Reviews using Efficient Repetitive Pre-Processing”, IEEE, January 2014.