International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 4, April2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

IMPLEMENTATION OF DATA MINING TECHNIQUES FOR CYBER CRIME DETECTION

K. Chitra Lekha*
Dr. S. Prakasam**

Abstract

Data mining technology is applied to fraud detection to ascertain the scam detection model, to depict the process of creating the scam detection model, and then to begin the data model with any classifier. As e-commerce transactions persist to develop, the allied online hoax remains an eye-catching resource of income for fraudsters. This counterfeit activity inflicts a significant financial hammering to merchants, making online fraud detection a prerequisite. The issue of scam detection is concerned with not only confining the fraudulent activities, but also detaining them as rapidly as possible. This relevance is decisive to shrink financial losses. Cyber crimes are a communal pest and rate our society greatly in numerous ways. The investigation of cyber crime cases has very significant role in police enforcement system in any country. This paper presents a comprehensive study on data mining techniques and its responsibility on detection of cyber crimes in real time applications. Data mining robotically sieves through massive quantity of data to uncover known/unknown patterns that fetches out valuable, innovative perceptions and formulate predictions. Data mining which is alienated into two learning skills viz., supervised and unsupervised is engaged to detect fraudulent asserts. Basically these techniques are used for fraud detection in many sectors such as health , insurance, E-commerce and it goes on.

Keywords:

Credit card scam;
Cyber bullying;
Genetic algorithm;
MultiLayer Perceptron;
Probability Density Evaluation;
Self Organizing Map;
Neural Networks.

Author correspondence:

K. Chitra Lekha,
Ph.D(Research Scholar)
SCSVMV University, Tamil Nadu, India.

* Ph.D (Research Scholar), SCSVMV University, Tamil Nadu, India.
** Associate Professor, SCSVMV University,Tamil Nadu, India.

International Journal of Engineering, Science and Mathematics
Vol. 7,Issue 4, April2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

1. Introduction

Data mining is the dominant technology to investigate the data proficiently from various perceptions and present them in to useful information. It is an authoritative upcoming technology with immense prospective to assist law enforcement so as to turn their attention on the most significant information in their crime databases. It utilizes machine learning, statistical and visualization skills to determine and forecast knowledge in a structure which is clear to the investigator. Data Mining techniques plays a crucial role like pulling out of requisite knowledge, discovering unsuspected information to make deliberate decision in a novel way which in term understandable by domain experts [1]. Data Mining is an analytical method intended to identify data in search of reliable patterns or systematic relations between variables, and then to authenticate the findings by applying the detected patterns to new subsets of data. The overall goal of data mining is to extract information from a dataset and transform it into useful structure for further use [2].

Fraud detection is one of the most complicated task not only in the aspect of technically, but also in crime investigations. The process of fraud detection is based on simple comparisons, but also based on association, clustering, perdition and outlier detection [3]. Crime is the task that troubles the public, augments the cruelty, tears down the possessions and contradicts the respect to nation. Cyber crime is all about the crimes in which communication channel and communication device has been used directly or indirectly as a medium whether it is a laptop, desktop, PDA, Mobile phones, watches, vehicle [4]. Cyber attacks may have some incentive behind it or may be processed accidentally. The attacks those are processed intentionally are termed as cyber crimes and they have severe collisions over the society in the structure of economical interrupt, emotional disorder, threat to National defense. Constraint of cyber crimes is relied on appropriate investigation of their deeds and understanding of their impacts over diverse levels of society [5]. The motive for cyber crime is the nature of the issues initiated from the information and communication technology remain more or less same across the world, however the economic, political and social conditions of all countries are different to each other [6]. Detection of cyber crimes is the recognition of indication of cyber crimes where no earlier incredulity exists. Initially it has to be learned that given data samples are deceptive or not. This can be done by learning either supervised or unsupervised. Supervised learning of cyber crime data set concerns with fake data that is formerly known and unsupervised learning of cyber crime data sets concerns with fake data that is not formerly considered as a fake data but after sometimes they imitate the nature of scam or crime. Then those data patterns are treated according to their deeds. Several terms are utilized for performing that task and they are depicted as techniques and methodologies for the detection of cyber crimes. Information security policies strengthen the security and well-being of information resource. They are the foundation and bottom line of information security with in the organization. Fraud remains a challenge for businesses and organizations in many fields. Data mining is an effective method for detecting various types of cyber crimes including telecommunication, credit card and medical insurance fraud as well as detecting intrusion to computer systems.

2. Data Mining Techniques – An Overview

2.1. Regression Analysis: Regression is a statistical methodology for exploring function with least error to model data and often utilized for numeric prediction. It is also employed in forecasting where its presence has significant overlap with the concept of machine learning. Its purpose is to discover the associations between independent and dependent attributes. The cautions are advisable because in some criteria it may also result in illusions or fake relationships among the attributes.

2.2. Association: Association aspires to discover fascinating association, correlations or informal patterns between the attributes in the databases. The discovered association rules can be

valuable for advertising products, product management, marketing goods etc. Using various assessments of patterns, well-built rules are explored. With the help of these explored strong rules, regularities among products in the data transactions recorded by Point-of-sale systems in provisional stores are identified. For example the rule {phishing, cyber stalking} {cyber crime} found in crime data of a state would that if a cyber criminal undergoes phishing and cyber stalking together, he is likely to perform credit card fraud or theft. Frequent pattern mining is a very huge researched study in the utilization of data mining with wide range of applications.

2.3. Clustering: Clustering recognizes the occurrences of similar cyber crimes cases that had taken place in a state. Once the clusters are determined, the items are tagged with their consequent clusters, and general patterns of the items in cluster are reviewed to outline a class description. For example, law enforcement may cluster the cyber criminals in to various sets based on the similarities of their age, residence, mode of committing crimes etc and form these general characteristics of the cyber criminals in a state can be utilized to depict those group of cyber criminals. This aids the law enforcement officials to recognize the cyber criminals better and to take necessary actions.

2.4. Classification: Classification is the source of model which concludes the class of an object depending on its attributes. Classification predicts definite constant valued function. For example, we can make classification model to catalog utilization of credit card as either valid or scam. A set of items is specified as training set in which every item is characterized by their vector of attributes along with its class. Thus the classification model for identifying the usage of credit card as valid or scam can be built by analyzing the relationship among the attributes. From the set of valid utilization of credit cards which serve as training set, a classification model can be constructed so that scam utilization of credit card can be determined.

3. Literature Review

B. Pushpalatha and C. Willson Joseph (2017) had examined several data mining techniques like Bayesian networks, Bayes minimum Risk, Genetic Algorithm, Hidden Markov model and Ontology and concluded that they had the potential to aid the detection of credit card frauds. Their findings had also highlighted that a learning strategy can provide enhanced fraud detection when it is used in conjunction with an established fraud detection system [7].

Atul Bamara and Mamta Bhatt (2013) had revealed the various cyber attack strategies adopted by cyber criminals to target the selected banks in Uttarkhand where spoofing, brute force attack , buffer overflow and cross side scripting are found positively correlated with Public and Private sector banks. Also their findings show a positive correlation between Intruder Detection and cyber attack that is online theft, hacking, malicious code, Dos attack and credit card/ ATM frauds as well as the above mentioned cyber attacks had positive correlation with System monitoring [8].

Raghavendra Patidar and Lokesh Sharma (2011) had tried to detect fraudulent transactions through the Neural network along with the Genetic algorithm. They used Genetic algorithm for making the decision about the network topology, number of hidden layers and number of nodes that would be used in the design of neural network of their credit card fraud detection. They also used artificial neural network for learning purpose which uses supervised learning feed forward back propagation algorithm [9].

Linda Delamaire, Hussein Abdou and John Pointon (2009) had recognized the diverse types of credit card deception; appraised optional techniques that have been utilized in scam detection. They briefed general terms in credit card deception and embarked key statistics and figures in that domain. They recommended that based upon the type of scam faced by banks or credit card companies, diverse measures can be adopted and implemented. Some of those measures include pair-wise matching, decision trees, clustering techniques, neural networks and genetic algorithms. Their proposals were designed to have beneficial attributes in terms of cost savings and time efficiency [10]**.**

K. Chitra Lekha and Dr. S. Prakasam (2017) had highlighted a common proposal about replica of Data Mining techniques and assorted cyber crimes in banking applications. They had renowned

patterns in scandalous manners in order to predict crime anticipate criminal commotion and thwart it. They had proposed a novel data mining techniques like K-Means, Influenced Association Classifier and J48 Prediction tree for the purpose of for scrutinizing the cyber crime data sets and thwarts out the reachable issues [11].

Paridhi Saxena and Anisha Malke (2014) had revealed that the reason for lack of proper statistics could be due to the fact that most of the existing laws and policies on information and technology does not mention anything regarding the cyber violence against women. They also stated that while India starting her journey in the field of Information technology, the priority was given to the protection of E-commerce and communication under IT Act 2000 whereas matters concerning cyber socialization and communications were not use their own methods to tackle such situations. But this, in no way helps to reduce or even prevent included. They had concluded that cyber crimes against women were basically the crimes against them with the motive of intentionally harming them and with the aid of modern telecommunication techniques like internet and mobile phones. Many women are either afraid or scared and many other the occurrence of cyber crimes [12].

Dr. Shalini Kashmiria (2014) had made an effort to emphasize cyber crimes against women in India which is a utterly new trend. The author had performed a proportional investigation amid the cyber laws that were adopting cyber crimes in India, United States of America and UK. The author had utilized secondary data source for collecting cyber crime data sets and adopted doctrinal as their research methodology. She had conferred about various cyber crimes against women in India and also affords apt implications for those crimes [13].

P. Rajesh and Dr. M. Suriakala (2016) had highlighted the effects of cyber stalking against women particularly that are psychologically weak. This research paper had implemented a quantitative and qualitative loom to scrutinize the internet allied tricks of women dwelling in diverse parts of Chennai city amid the age of 17 to 40 years. They had also fetched out the procedures that generated alertness on the secure and perilous practice of handling their online space. Their research had revealed diverse sorts of defensive procedures to be griped by women on their cyber space to defend themselves from cyber stalking [14].

Aarti Bansal (2015) had explored the concert of major data mining techniques that is Apriori, Decision trees and K- Nearest Neighbor for investigating crimes against women. She had realized the vital requirement to scrutinize the crime data and had developed the tools that can assist the concerned authorities to take appropriate actions to alleviate escalating crime against women. The concert is précised in terms of time taken to build the model, classification of instances in to correctly and incorrectly instances and precision [15]

Jacqulin Margret J and Shrijina Sreenivasan (2013) Jacqulin Margret J and Shrijina Sreenivasan (2013) had spotlighted upon perceiving fake medical symphony and had demonstrated the framework of applying data mining to the medical databases so that the given medical data can be sorted as real or false with the reasonable precision. The relevance of data mining techniques to medical scam recognition systems would facilitate in discovering patterns that would perceive incongruity in medical data. The mined medical data undergoes data preprocessing so that the overall effect of such a realization is high eminence data that is potentially more precise and which thwart incidence of scam in case of forged medical trade and receipts. They had highlighted that this procedure can be utilized to perceive the trade of potentially perilous drugs by pharmacists thereby thwarting such medical scam [16].

Pravin R. Bagde and Manoj S. Chaudhari (2016) had developed a novel hybrid loom by merging the merits of supervised and unsupervised data mining techniques for the purpose of investigating deceptive assets in health insurance industry. This hybrid approach of support vector machine and Bisecting K-Means Clustering Method are utilized to perceive and evade the scam in health insurance province. In order to recognize scam from bulk of medical data , they had utilized statistical data mining techniques [17].

Hossein Joudaki, Arash Rashidian and Mohammad Arab (2015) had evaluated studies that execute data mining techniques for perceiving health care hoax and cruelty, by utilizing supervised and unsupervised data mining looms. They had recognized that most existing studies had focused on algorithmic data mining without an accent on or claim to hoax recognition efforts in the perspective of vigor service prerequisite or health insurance policy. They had concluded that more studies are required to unite sound and evidence-based verdict and cure looms towards the counterfeit or abusive behaviors [18].

4. Data Mining Techniques for detecting Cyber Crimes

4.1 Detecting Cyber crimes in Banking Sector: Data mining technique assists to evaluate patterns and dealings that escort to scam. Fraud supervision is a knowledge-intensive task. The most important aspect in scam detection is to forecast the genuine transactions performed by the account holders. For this purpose, Data mining techniques are utilized which ensures and separates the transactions that do not fit in to a precise group or not regular replicate transactions.

4.1.1. The Clustering technique in detecting cyber crime: Clustering aids in consortium of the data into related cluster that assists in simple retrieval of data. Clustering technique splits the data in to interconnected constituents in such a mode that patterns and orders becomes detectable. This approach makes the use of constraints' data clusterisation sections.

4.1.2. Probability density evaluation approach: Gaussian mixture replica is utilized to model the probability density function. The latter is the summation of weighted constituent densities of Gaussian form. This approach carried out a statistical representation of past deeds and outputs the novelty appraisal of present custom as a negative log likelihood of recent usage. The detection decision is then concluded from the output of this novelty filter.

4.2. Detecting Cyber crimes against Women

4.2.1. Association Rule Mining in detecting cyber crime patterns: Exploring recurrent cyber crime patterns assists in recognizing pattern that emerges repeatedly in cyber crime datasets. In numerous aspects of cyber crime data mining techniques such as Association rule mining, classification, clustering, correlation etc, frequent item set mining emerges as an significant task. The PredictiveApriori technique digs out the cyber crime patterns that take place frequently in cyber crime dataset against women. Thus it can be utilized in the renovation of quantitative data in to qualitative data.

4.2.2. Deduction of Cyber bullying against women: Cyber Bullying is an emerging issue in the social media and it is fetching to be chief cyber threat to teenage girls and women. For the purpose of detecting cyber bullying threats against women, the methodology of Text mining can be performed such as online sexual predator identification and spam exposure. The character of user whether he is a bully or sufferer can be guessed when the data uploaded or text messaged by him is analyzed by the concept of Text analysis. With the combined utilization of Text mining and data mining, the conclusion is driven out whether there exists cyber bullying against women exists or not by investigating the text contents posted in the social media.

4.3. Detecting Cyber crimes in Health care

4.3.1. MultiLayerPerceptron Neural network for detecting health care fraud: This technique is utilized in detecting frauds in medical field since it can handle large data structures particularly non-linear associations and it has high tolerance to discrepancy data. In order to decrease the discrepancies of experts' classifications, this technique classifies the prescription profiles of universal practitioners. The probabilistic elucidation was utilized to filter the classified profile behind every training of the network. The low-probability filtered profile that is incorrectly classified profiles were recognized and then evaluated by the expert consultants.

4.3.2. Decision Tree approach: Decision tree approach is chosen since its generated rules and outcomes are easily understood. C5.0 decision tree classifier utilizes the pruning technique which permits tuning the severity of tree. The adaptive boosting in C5.0 constructs a series of classifiers

and utilizes a selection approach to attain the ultimate classification, and misclassification weights describing various costs for diverse errors in classification.

4.4. Detecting Cyber crimes in E- Commerce The usage of credit cards had grown up due to the development of E-Commerce technology. At present, with the increase in the credit card transactions, the credit card frauds had become a common cyber threat as the credit card became the vital form of payment. To retain the consistency of the payment scheme, it is essential to build an enhanced fraud detection system in business organizations. As in E- Bank, many transactions endure concurrently, so a fraud detection system should distinguish between valid, suspicious fraud and an illegal transaction.

4.4.1. Genetic Algorithm for the detection of Credit Card scam: The Genetic algorithms tends to attain the enhanced patterns to precisely eradicate the fraud and also to builds a secured and well-organized e-payment scheme to identify whether a transaction is fake or not. Based on the customer deeds, the Genetic algorithm has to deduct scams and minimize the number of false alerts while undergoing credit card transactions. The best elucidation is driven out by repeating a pre-specified number of iterations. The attributes required to generate fraud transactions include current bank balance, Credit card over draft, Credit card custom frequency count and location etc. The goal is to achieve the enhanced and best possible outcomes. As soon as credit card transactions are performed in the E-Banks, the Genetic algorithm predicts the probability of fraud transaction and implements a sequence of anti-fraud policies to thwart banks from huge losses and shrink cyber threats.

4.4.2. Self-Organizing Map neural network for the detection of Credit Card scam: The purpose of SOM is to develop client profile and evaluate fraud patterns in the detecting the credit card scam. Initially it recognizes the credit card data transactions; preprocesses them and provide them as input to the SOM and it iteratively fine-tunes the weights of neurons so that the data had been classified in to valid and fake sets. The purpose of input and mapping layers is to categorize, group, discover and obtain hidden cyber crime patterns in the input data and proceed as a filtering method for auxiliary layers. In the payment system, this algorithm categorizes all credit card transactions in to valid and fake sets by pursuing two hypotheses: In this technique all transactions in the payment system are classified into valid and fake sets by following the two hypotheses:

1. If a fresh received transaction is analogous to all prior transactions from valid set, then it is regarded as valid credit card transaction.
2. If a fresh received transaction is analogous to all prior transactions from the fake set, then it is regarded as scam.

5. Conclusion

Perceiving and thwarting cyber crimes is tricky, because cyber criminals innovate latest proposals all the time, and those schemes develop more and more sophisticated to elude easy detection. In meticulous, understanding the correlation among analysis competence and the characteristics of cyber crime type can facilitate investigators more effectively utilize those techniques to recognize trends and patterns, tackle problem areas and even predict forthcoming cyber crimes. The detection process should be adjustable to allow the system to deal with the constantly changing nature of crimes. Similarity measures are an important factor which helps to find unsolved crimes in crime pattern. This research paper has presented only a selection of the various data mining techniques for cyber crime detection in different fields. Some of those techniques have persisted and proven to be successful, while others are in the process of development and enhancement to better apply to new fraudulent acts. After all, it is not the organization alone who suffers from the consequences of fraud, but all the individuals and stakeholders related to that organization will be victims. Therefore, organizations are entirely accountable for learning the best practices and choosing the best method that matches their needs in order to safeguard against cyber crimes.

References

[1] R. Jayabrabu, Dr. V. Saravanan, Prof. K. Vivekanandan, "A Framework: Cluster Detection and Multidimensional Visualization of Automated Data Mining Using Intelligent Agents", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.1, January 2012.

[2] Karan Pruthi and Dr. Prateek Bhatia, "Application of Data Mining in Predicting Placement of Students", IEEE, 2015.

[3] Dr. R. Jayabrabu, Dr. V. Saravanan and Dr. J. Jebamalar Tamilselvi, "A Framework for Fraud Detection System in Automated Data mining using Intelligent agent for better Decision making process".

[4] Vinit Kumar Gunjan, Amit Kumar and Sharda Avdhanam,"A survey of Cyber crime in India", IEEE, 2013.

[5] Hemraj Saini, Yerra Shankar Rao and T.C. Panda, "Cyber- Crimes and their Impacts: A Review", International journal of Engineering Research and Applications (IJERA)", Vol.2, No.2, March-April 2012.

[6] Shobana Jeet, "Cyber crimes against women in India: Information Technology Act, 2000", Elixir, 2012.

[7] B. Pushpalatha and C. Willson Joseph, " Credit Card Fraud Detection based on the Transaction by using Data mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering(IJIRCCE), Vol.5, No.2, February 2017.

[8] Atul Bamrara, Gajendra Singh and Mamta Bhati, "Cyber attacks and Defence strategies in India: An Emprical Assessment of Banking Sector", International Journal of Cyber Criminology, Vol. 7, No.1, January-June 2013.

[9] Raghavendra Patidar and Lokesh Sharma, "Credit Card Fraud Detection using Neural Network", International Journal of soft Computing and Engineering, Vol. 1, No. NCA12011, June 2011.

[10] Linda Delamaire, Hussein abdou and John Pointon, "Credit card fraud and Detection Techniques: A Review", Banks and Bank Systems, Vol. 4, No. 4, 2009.

[11] K. Chitra Lekha and Dr. S. Prakasam, "Data mining Techniques in Detecting and Predicting Cyber crimes in Banking Sector", IEEE- International Conference on Energy, Communication , Data Analytics and Soft Computing, No. 3, August 2017.

[12] Paridhi Saxena and Anisha Malke, "Cyber Crimes: Another Dimension of Women Victimization", International Journal of Research and Analysis, Vol. 2, No. 3, 2014.

[13] Dr.Shalilni kashmiria, "Mapping Cyber crimes against Women in India", International Research Journal of Commerce and Law, Vol. 1, No. 5, December 2014.

[14] P. Rajesh and Dr. M. Suriakala, "An Analytical study on Cyber Stalking awareness among Women using Data mining Techniques", Journal of Research in Computer Science, Engineering and Technology, Vol. 2, No. 3, September 2016.

[15] Aarti Bansal, "Performance Comparison of Data Mining Techniques to analyze crime against Women", International Journal of Scientific Research and Education, Vol. 3, No. 9, October 2015.

[16] Jacqulin Margret J and Shrijina Sreenivasan, "Implementation of Data mining in Medical fraud Detection", International Journal of Computer Applications, Vol. 69, No. 5, May 2013.

[17] Pravin R. Bagde and Manoj S. Chaudhari, "Improving Fraud Detection Mechanism in Health Insurance Industry using Data mining and Statistical Techniques", International Journal for Scientific Research & Development, Vol. 4 No. 5, 2016.

[18] Hossein Joudaki, Arash Rashidian and Mohammad Arab, "Using data mining to detect Health care fraud and Abuse : A Review of Literature", Global Journal of Health Science, Vol. 7, No. 1, 2015.