JCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Volume-4, Issue-5 E-ISSN: 2347-2693

A Novel Approach for Detection of Fraud Using SOM

Swati Sucharita Barik

Department of Computer Science & Engineering,
Centurion University of Technology and Management, Odisha, India

Available online at: www.ijcseonline.org

Received: May/03/2016 Revised: May/14/2016 Accepted: May/22/2016 Published: May/31/2016

Abstract— Fraud refers to abuse of organization’s system illegally. Fraud is the crime or offense of deliberately deceiving in order to damage them to obtain property or services unjustly. Fraud can be accomplished through the aid of forged objects. The scope of this paper is to investigate whether data mining techniques can be used for detecting fraud.

Keywords—Fraud, Data Mining,Neural Computing,Self Organizing Map

I. INTRODUCTION

The Association of certified fraud examiners (ACFE) defined fraud as “the use of one’s business for personal gain through the deliberate misuse of employing the assets of the organization."[1].Fraud can be defined as an illegitimate activity by a person other than an eligible person.The Oxford Dictionary defined fraud as criminal deception.Through the mechanism, the fraudster gets access to unlawful advantage by causing unlawful loss. A dishonest person may be called a fraud.Fraud commonly occurs while in buying or selling of property.

II. TYPES OF FRAUD

The **Internet Fraud** refers to the type of fraud,which uses online services.Online Services include chat rooms,e-mails,websites to present fraudulent solicitations to victims,to make fraudulent transactions.Internet Fraud is committed in different ways.According to an analysis by FBI,U.S Companies’ losses due to internet fraud in the year 2003 passed upto US$500 million.

Internet serves as an excellent tool for the investors as well as the fraudsters.It provides easy,inexpensive investment opportunities.

Mail Fraud refers to an approach which obtains money or other valuable objects unlawfully.Many times the postal system can be used for making the criminal offence.

Cellular Fraud(Cloning) is a type of fraud,where the fraudulent usage is imposed on the legitimate use of an account.It causes difficulty to customers and the service providers.

Nigerian Letter Fraud is a kind of fraud,where a govt. official attempted to transfer money to US.but in rel therewas no money at all.Many became the victims of it and million of dollars had been lost.Some victims have been lured to Nigeria,the place where they have been imprisoned.

Identity Theft occurs when a person assumes another’s identity to make some illegal activity. For example, informations have been get from sources like one’s wallet, credit card information,bank information. Then after fraudster will gather information to make fraud financially.

Spyware is a kind of software which secretly collects the personal information on internet.

Adware tracks visitor’s interest on internet,monitors the types of sites visited and then the informations are sold to a third party for marketing purpose.

Advance Fee Scheme requires the victim to pay some money anticipating a large amount of money to get.

Fraudsters offer some financing opportunities to victims.

Internet fraud or **wire fraud**,which is also known as **mail fraud,** refers to a type of fraud scheme which uses one or more online services like chat rooms,emails,websites to conduct fraudulent transactions.

Friendly Fraud describes a customer making internet purchase with its credit card and then issues a charge back through card provider after receiving goods.

Credit Card Fraud is a growing problem in today’s era.The fraud occurs due to lost and stolen cards. The credit card and debit card fraud are considered as a crime where these cards are reproduced by fraudsters. This type of crime is known as **'skimming'**. Credit or debit card fraud can also occur when the card is lost or stolen and used by fraudster to purchase goods or remove cash from ATMs or other locations. The Confederation of British Industry reports as of 2001, two thirds of UK businesses have experienced a serious online incident like hacking, virus attacks or credit card fraud. For these reasons the forms of credit card fraud counterfeit, i.e card not present and lost and stolen cards.

III. FRAUD DETECTION

The advanced transactional behavioral analytics are the primary key to detection process[2]. These systems analyze card transactions, spot out-of-character spending by cardholders and the patterns of behavior indicative of fraud. The banks are using this technology which has been so successful over the past decade at reducing credit card fraud loss rates to debit cards too.PINs,which fraudsters are skilled at obtaining, are no longer considered sufficient to keep these accounts secure. Now a days proposed systems are effective at catching criminal organizations,by detecting no. of sequences of fraud after the no. of transactions. Advanced analytics like neural network models are used to examine hundreds of nonlinear data points.

Phishing is a practice of trying fraudulently to get consumer banking and credit card information.It is a new problem occurred for online banking[3].

In many cases, the fraudster place a fraudulent order with a valid issued and active credit card number.Though the card processors request additional information for expiration date, AVS (Address Verification Service) and CVM (Card Verification Method) checks, these fields are not mandatory and do not result in declined transactions.Fraudsters can gain access to valid credit card numbers in a number of ways. The credit card numbers are rarely stolen in cyberspace.Today all online merchants use secure communication channels like secure socket layer, or SSL,when important data is transmitted between the consumer browser and the web site, therefore the fraudster cannot intercept card numbers during a transaction.A greater risk is theft of credit card data from storage on the merchant's web site.These risks can be addressed by implementation of appropriate site security.Stolen or lost credit cards provide fraudsters the full access to account information,expiration data and billing name.These card numbers provide opportunity to the fraudster since the legitimate cardholder will report the incident to the issuer and the account will be blocked.Credit card numbers collected from card imprints, receipts or monthly statements collected in dumpsters give fraudsters a wider window of opportunity, since the cardholder is unaware that the card number has been compromised until he or she receives a statement from the issuer that includes unauthorized transactions.New high tech tools commonly used to steal credit card information are hand-held credit card skimmers. These devices can read the card information encoded in the magnetic stripe and store thousands of card numbers that are later uploaded to a PC. Since these devices are easily concealed, an unethical waiter can easily swipe the card while walking between the cash register and the table. As with credit card numbers stolen from imprints and receipts, the cardholder is typically unaware of the event for weeks or even months.

Card issuers first established the Address Verification System (AVS) as a security mechanism for card-not-present transactions.AVS validates the billing address information provided by the consumer against the billing address information that the issuer has on record for the account. AVS checks the ZIP code and the numeric part of the street address and returns a match/mismatch response.The AVS response provides additional information for the merchant, but AVS match is not required for approval,nor it is a transaction,obtained AVS match response guaranteed against chargebacks.The decision on whether to accept an order based on the AVS response is left to the merchant. Fraud detection systems are widely used in telecommunications, online transactions, the insurance industry,computer and network security.Effective fraud detection systems use both fraud rules and pattern analysis. Fraud detection is a continuously evolving process.Many users committing fraudulent behavior are not aware of the fraud detection methods which have been successful in the past and will adopt strategies leading to identifiable frauds. The earlier detection tools need to be applied as well as the latest developments.The main purpose of fraud detection systemsis to identify general trends of suspicious or fraudulent applications and transactions. In the case of application fraud, these fraudsters apply for insurance entitlements using falsified information, and apply for credit and telecommunications products/services using non-existent identity information or someone else's identity information. In case of transactional fraud, these fraudsters take over or add to the usage of an existing legitimate credit or telecommunications account.Fraud detection is related to intrusion detection, a field of computer security detecting attacks on computers and computer networks.

IV. FLOW OF DATA MINING PROCESS

Fraud detection tools are composed of a variety of analysis approaches.Supervised learning algorithms samples of both fraudulent and non-fraudulent records are used to construct models which allow to assign new observations into one of the two classes that produce classifiers using rules of the "if x, then y" form.Examples of such algorithms include BAYES, RIPPER,Tree-based algorithms such as CART. Neural network techniques have been most commonly used for fraud detection.But not only neural networks but also various data mining methods were used to model.

Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions[4]. The first and simplest analytical step in data mining is to describe the data, summarize its statistical attributes.

Data mining finds patterns and relationships in data by using sophisticated techniques to build models,abstract representations of reality.Data mining is the process of discovering previously unknown relationships in collections of structured data using machine learning and statistical analysis techniques.Creating a data mining solution is the practical application of data mining. The solution may combine database management,data warehouses, text mining tools to structure the data,data mining software tools, data visualization technology and advanced data analysis.The best data mining solutions are defined by each client's unique business,organizational,and technology environment. One can mine its data wherever they reside, in a data warehouse, database, legacy system, or even in external information such as survey and purchased data sets.

Common applications include:

- Fraud and abuse reduction

- Network optimization and intrusion detection
- Web usage analysis
- Customer service
- Call center analysis
- Human genome analysis
- Insurance claims processing analysis
- Resource management
- Accident trending
- Crime pattern detection
- Targeted marketing
- Disease outbreak detection

The sub components of the data mining as follows:

1. Data gathering and preprocessing.

2.Selecting important attributes i.e feature extraction.

3.Modeling fraud detection mechanism by previous fraudsters' activity data sets.

4. Training and testing phase of the data mining process.

5.Obtaining classification rules for knowledge discovery.

V.OCCURENCE OF FRAUD AND SOM

Ukraine tops the list with staggering 19% fraud rate closely followed by Indonesia at 18.3% fraud rate.In the list of high risk countries are Yugoslavia (17.8%), Turkey (9%) and Malaysia (5.9%). Surprisingly United States, with its high number Credit Card transactions, has a minimum fraud rate. Over the last few years, the credit card industry in UK was subjected to maximum threat from increasing fraud losses[5].

The following table shows the trend in volumes of frauds.

Method	Percentage
Lost or stolen cards	48%
Identity theft	15%
Skimming or cloning	14%
Counterfeit card	12%

(Fig:1- Method of card fraud and their percentage of occurence)

Data mining software analyzes relationships and patterns in stored transaction data based on open ended user queries.Several types of analytical software are available, statistical, machine learning, and neural networks.

Generally, following four types of relationships are sought,

•Classes:Stored data is used to locate data in predetermined groups.For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

•Clusters: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.

•Associations:Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

•Sequential patterns: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Types of Analysis:

•Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.

•Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.

•Decision trees: Tree-shaped structures that represent sets of decisions.These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that can be applied to a new unclassified dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.

•Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k 1). Sometimes called the k-nearest neighbor technique.

•Rule induction: The extraction of useful if-then rules from data based on statistical significance.

•Data visualization: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

•Self organizing map

Self-organizing maps(SOMs) are data visualization technique invented by Professor Teuvo Kohonen which reduces the dimensions of data through the use of self-organizing neural networks[8][14].The problem that data visualization attempts to solve is that humans simply cannot visualize high dimensional data as is so techniques are created to help us understand this high dimensional data. Two other techniques of reducing the dimensions of data that has been presented in this course has been N-Land and Multi-dimensional Scaling. The way SOMs go about reducing dimensions is by producing a map of usually 1 or 2 dimensions which plot the similarities of the data by grouping similar data items together. So SOMs accomplish two things, they reduce dimensions and display similarities. The SOM is an algorithm used to visualize and interpret large high-dimensional data sets. Typical applications are visualization of process states or financial results by representing the central dependencies within the data on the map. The Self-Organizing Map belongs to the class of unsupervised and competitive learning algorithms. It is a sheet-like neural network, with nodes arranged as a regular, usually two dimensional grid. As explained in the previous section on Neural Networks, we usually think of the node connections as being associated with a vector of weights. In the case of Self-Organizing Maps, it is easier to think of each node as being directly associated with a weight vector.

SOM is a neural network with feed-forward topology and an unsupervised training algorithm that uses a self-organizing process to configure its output neurons according to the topological structure of the input data. The self-organizing process is based on competitive training and consists in tuning the weights.

The Self-Organizing Map belongs to the class of unsupervised and competitive learning.

VI. PROPOSED WORK,APPROACH WITH SIMULATION RESULTS

SOM is being employed into the fraud detection.It is employed to differentiate the transaction input data into two sets i.e. genuine set and fraudulent set.

For the said detection of fraud two hypotheses are made into consideration. They are as follows:

1. If the transaction made just is similarly equal to the transactions made previously in genuine set, then it is treated as a genuine one.

2. If the transaction made just is similarly equal to the transactions made previously in fraudulent set, then it is treated as a fraudulent one.

So the main idea behind it is to formulate legal card holder and fraudster.

APPROACH: The Approach for the fraud detection consists of authentication and screening layers,risk scoring and behaviour analysis layer.it contains a layer of SOM followed by a feed forward neural network or a rule based risk scoring method.

The SOM layer has three purposes:

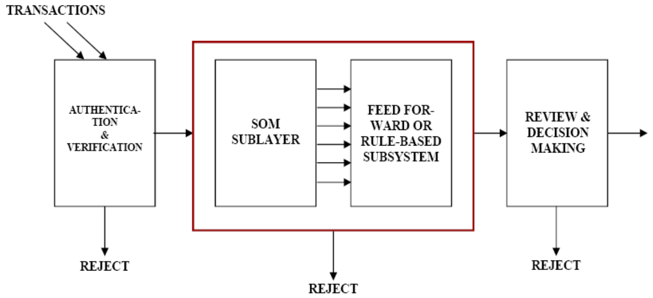
-To classify and cluster the data.

- To detect and derive hidden patterns in the input data.

- To act as a filtering mechanism for further layers.

Layman, merchants and banks do not have idea about fraud case. When data is fed to SOM,many clusters can be emerged.

SOM need not to classify the fraudulent and genuineness of data. But it can be helpful to classify data like safe, rarely safe, fraud prone etc.the sub layer of feed forward neural network or rule based system can be able to process and analyse output. It is mainly used for getting trends of fraud and its risk.


TRANSACTIONS
AUTHENTICA-TION & VERIFICATION
SOM SUBLAYER
FEED FOR-WARD OR RULE-BASED SUBSYSTEM
REVIEW & DECISION MAKING
REJECT
REJECT
REJECT

(Fig:2-SOM Layers)

The input to the SOM sub-layer is in the form of multidimensional vectors of real-world transaction attributes, which can be broadly classified as: Customer-related, Account-related and Transaction-related. Most of customer and account related attributes are static and are stored in databases of banking systems. We created tables with appropriate records for customers and their accounts. We considered transactional data of a banking database with data input.

Normalizing of values can be made by $Z = (Xi - \mu)/\sigma$,which gives out the normal distribution.

Calculating the mean and std. deviation from the transaction amts,

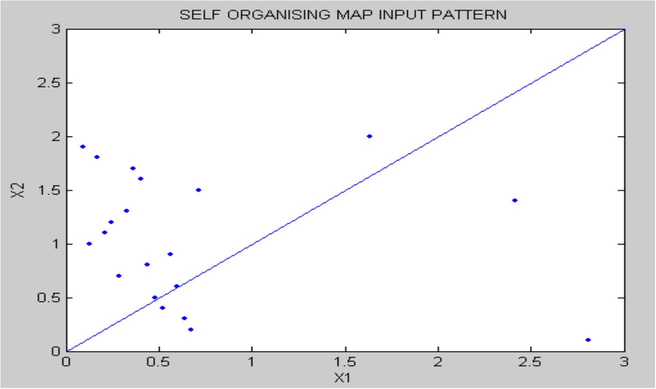
μ =14150 and

σ=12771.8

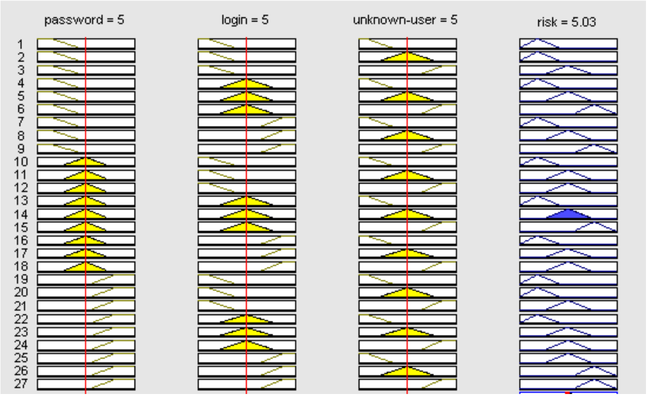
Using μ and σ ,we can get the normalized values for each transaction.(20 values are taken).

Taking the initial learning rate α = 0.9, the SOM algorithm is expected to generate a weight matrix and group the original weights by computing the Euclidean distance between all the normalized inputs.

The output can be interpreted and used in various ways so as to arrive at important results: The clustering of data into dense and sparse clusters shows the categories of transactions performed more frequently and rarely by each customer.



(Fig:3-Simulation Result For Normalized Inputs ‘X’)



(Fig:4-All possible rules and average risk)

CONCLUSIONS:Data mining tools and techniques are widely used for the problem of fraud. Internet platform is a huge place with a no of customers, merchants, banks etc.These customers will reduce their online spending due to the frauds arising in a huge,commited by the fraudsters. So the only way to detect the fraud is by the dynamic way, such that frauds can be reduced.

ACKNOWLEDGMENT

I would like to take blessings from The almighty, my parents, my Family,friends,Elders for their willingness to spare and contribute their precious time, guidance, sharing their knowledge,support and kind cooperation.

REFERENCES

[1] S. Ghosh and D.L Reilly,”Credit Card fraud detection with a neural network.”in proc. Of 27th Hawaii Int. conf. syst. Sci.,pp.621-630.

[2] K. Fanning ; K. O. Cogger ; R. Srivastava,”Detection of Management Fraud”, Publication Year: 1995, Page(s):220-223

[3] G. C. Y. B. N. Grozavu, "Unsupervised Learning for Analyzing the Dynamic Behavior of Online Banking Fraud," International Conference on Data Mining Workshops IEEE, 2013.

[4] Using data mining to detect fraud.SPSS technical report.2000.

[5] Fuzail Misarwala, KausarMukadam, and Kiran Bhowmick, "Applications of Data Mining in Fraud Detection", International Journal of Computer Sciences and Engineering, Volume-03, Issue-11, Page No (45-53), Nov -2015, E-ISSN: 2347-2693.

[6] Philip K. Chan,Wei Fan and J Stolfo,1999,IEEE Distributed Data Mining in credit card fraud detection.

[7] L. A. V. D. A. M. d. C. F. S. M. Emanuel Mineda Carneiro, "Cluster Analysis and Artificial Neural Networks : A Case Study in Credit Card Fraud Detection," International Conference on Information Technology - New Generations IEEE, 2015.

[8] Teuvo Kohonen,”The Self-Organizing Map", Proceedings of the IEEE 78, no.9(sept 1990).

[9] Fanning, K. and K. Cogger (1998). Neural network detection of management fraud using published financial data. International Journal of Intelligent Systems in Accounting, Finance & Management 7, 21-41.

[10] Jiwei Han & Kamber,“Data Mining Concepts and Techniques".

[11] Yufeng Kou,Chang-Tien lu,Sirirat Sinvongwattana Yo-ping Huang,”Survey Of Fraud Detection Techniques "Proceedings Of 2004 IEEE Intern. Conf. On networking,Taipei,March 21-23,2004.

[12] K. Chikin and I. Shlyik, “Countering Illegal Transactions in Internet Purchasing Systems,” World of Cards 7 (2002): 15-21.

[13] Bhatla, Prabhu, and Dua, “Understanding Credit Card Frauds.”

[14] T. Kohonen, "An introduction to neural networks", Neural Networks, vol. 1, pp. 3-16, 1988

AUTHOR PROFILE

Swati Suchrita Barik is currently working as an Asst. Professor in CUTM,Bhubaneswar,Odisha,.She holds M.Tech Degree in Computer Science & Engineering.She is having 9 years of experience.Her area of research interests include computational intelligence,artificial intelligence,Data mining and cryptography & network security.