International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 4, Number 13 (2014), pp. 1343-1348

http://www. irphouse.com

Credit Card Fraud Detection Using Self Organised Map

Mitali Bansal and Suman

C. S. E. Dept.,
Hindu College of Engineering, Sonepat, Haryana, India

Abstract

Self organizing Maps (SOMs) are most well-known, unsupervised approach of neural network that is used for clustering and are very efficient in handling large and high dimensional dataset. As SOMs can be applied on large complex set, so it can be implemented to detect credit card fraud. Online banking and e-commerce has been experiencing rapid growth over past years and will show tremendous growth even in future. So, it is very necessary to keep an eye on fraudsters and find out some ways to depreciate the rate of frauds. This paper focuses on Real Time Credit Card Fraud Detection and presents a new and innovative approach to detect the fraud by the help of SOM.

Keywords— Self-Organizing Map, Unsupervised Learning, Transaction

Introduction

The fast and rapid growth in the credit card issuers, online merchants and card users have made them very conscious about the online frauds. Card users just want to make safe transactions while purchasing their goods and on the other hand, banks want to differentiate the legitimate as well as fraudulent users. The merchants that is mostly affected as they do not have any kind of evidence like Digital Signature wants to sell their goods only to the legitimate users to make profit and want to use a great secure system that avoid them from a great loss. Our approach of Self Organizing map can work in the large complex datasets and can cluster even unaware datasets. It is an unsupervised neural network that works even in the absence of an external teacher and provides fruitful results in detecting credit card frauds.

It is interesting to note that credit card fraud affect owner the least and merchant the most. The existing legislation and card holder protection policies as well as insurance scheme affect most the merchant and customer the least. Card issuer bank also has to pay the administrative cost and infrastructure cost. Studies show that average time lag between the fraudulent transaction dates and charge back notification

can be high as 72 days, thereby giving fraudster sufficient time to cause severe damage.

In this paper first, you will see a brief survey of different approaches on credit card fraud detection systems,. In Section 2 we explain the design and architecture of SOM to detect Credit Card Fraud. Section 3, will represent results. Finally, Conclusion are presented in Section 4.

A Survey of Credit card fraud Detection

Fraud Detection Systems work by trying to identify anomalies in an environment [1].

At the early stage, the research focus lies in using rule based expert systems. The model's rule constructed through the input of many fraud experts within the bank [2]. But when their processing is encountered, their output become was worst. Because the rule based expert system totally lies on the prior information of the data set that is generally not available easily in the case of credit card frauds. After these many Artificial Neural Network (ANN) is mostly used and solved very complex problems in a very efficient way [3].

Some believe that unsupervised methods are best to detect credit card frauds because these methods work well even in absence of external teacher. While supervised methods are based on prior data knowledge and surely needs an external teacher. Unsupervised method is used [4] [5] to detect some kind of anomalies like fraud. They do not cluster the data but provides a ranking on the list of all segments and by this ranking method they provide how much a segment is anomalous as compare to the whole data sets or other segments [6].

Dempster-Shafer Theory [1] is able to detect anomalous data. They did an experiment to detect infected E-mails by the help of D-S theory. As this theory can also be helpful because in this modern era all the new card information is sent through e-mails by the banks. Some various other approaches have also been used to detect Credit Card Frauds, one of which is ID3 pre pruning method in which decision tree is formed to detect anomalous data [7].

Artificial Neural Networks are other efficient and intelligent methods to detect credit card fraud. A compound method that is based on rule-based systems and ANN is used to detect Credit card fraud by Brause et al. [8].

Our work is based on self-organizing map that is based on unsupervised approach to detect Credit Card Fraud. We focus on to detect anomalous data by making clusters so that legitimate and fraudulent transactions can be differentiated. Collection of data and its pre-processing is also explained by giving example in fraud detection.

SYSTEM DESIGN ARCHITECTURE

The SOM works well in detecting Credit Card Fraud and all its interesting properties we have already discussed. Here we provide some detailed prototype and working of SOM in fraud detection.

Our Approach to detect Credit Card Fraud Using SOM

Our approach towards Real time Credit Card Fraud detection is modelled by prototype. It is a multilayered approach as:

1. **1.** Initial Selection of data set.
2. **2.** Conversion of data from Symbolic to Numerical Data Set.
3. **3.** Implementation of SOM.
4. **4.** A layer of further review and decision making.

This multilayered approach works well in the detection of Credit Card Fraud. As this approach is based on SOM, so finally it will cluster the data into fraudulent and genuine sets. By further review the sets can be analyzed and proper decision can be taken based on those results. The algorithm that is implemented to detect credit card fraud using Self Organizing Map is represented in Figure 1:

1. Initially choose all neurons (weight vectors w_i) randomly.
2. For each input vector I_i
   {
   2. 1) Convert all the symbolic input to the Numerical input by applying some mean and standard deviation formulas.
   2. 2) Perform the initial authentication process like verification of Pin, Address, expiry date etc.
   }
3. Choose the learning rate parameter randomly for eg. 0. 5
4. Initially update all neurons for each input vector I_i.
5. Apply the unsupervised approach to distinguish the transaction into fraudulent and non-fraudulent cluster.
   5. 1) Perform iteration till a specific cluster is not formed for a input vector.
6. By applying SOM we can divide the transactions into fraudulent (F_k) and genuine vector (G_k).
7. Perform a manually review decision.
8. Get your optimized result.

Figure 1: Algorithm to detect Credit Card Fraud

Initial Selection of Data Set

Input vectors are generally in the form of High Dimensional Real world quantities which will be fed to a neuron matrix. These quantities are generally divided as [9]:

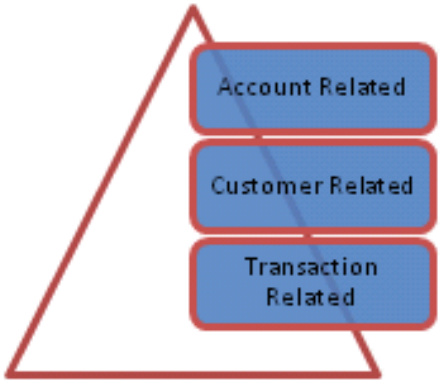


Figure 2: Division of Transactions to form an Input Matrix

In Account related quantities we can include like account number, currency of account, account opening date, last date of credit or debit available balance etc. In customer related quantities we can include customer id, customer type like high profile, low profile etc. In transaction related quantities we can have transaction no, location, currency, its timestamp etc.

Conversion of Symbolic data into Numeric

In credit card fraud detection, all of the data of banking transactions will be in the form of the symbolic, so there is a need to convert that symbolic data into numeric one. For example location, name, customer id etc. Conversion of all this data needs some normal distribution mechanism on the basis of frequency. The normalizing of data is done using $Z = (N_i - M_i) / S$ where N_i is frequency of occurrence of a particular entity, M is mean and S is standard deviation. Then after all this procedure we will arrive at normalized values [9].

Implementation of SOM

After getting all the normalized values, we make a input vector matrix. After that randomly weight vector is selected, this is generally termed as Neuron matrix. Dimension of this neuron matrix will be same as input vector matrix. A randomly learning parameter α is also taken. The value of this learning parameter is a small positive value that can be adjusted according to the process. The commonly used similarity matrix is the Euclidian distance given by equation 1:

Distance between two neuron =

$$j_x(p) = \min_j \| X - W_j(p) \| = \{ X_i - W_{ij}(p)]^2 \}^{1/2}, \quad (1)$$

Where j=1, 2……m and W is neuron or weight matrix, X is Input vectorThe main output of SOM is the patterns and cluster it has given as output vector. The cluster in credit card fraud detection will be in the form of fraudulent and genuine set represented as F_k and G_k respectively.

Review and decision making

The clustering of input data into fraudulent and genuine set shows the categories of transactions performed as well as rarely performed more frequently as well as rarely by each customer. Since by the help of SOM relationship as well as hidden patterns is unearthed, we get more accuracy in our results. If the extent of suspicious activity exceeds a certain threshold value that transaction can be sent for review. So, it reduces overall processing time and complexity.

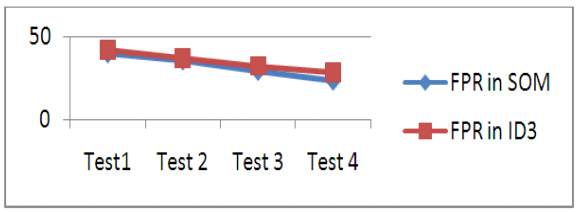
Results

The no of transactions taken in Test1, Test2, Test3 and Test4 are 500, 1000, 1500 and 2000 respectively. When compared to ID3 algorithm our approach presents much efficient result as shown in figure 3.

Conclusion

As results shows that SOM gives better results in case of detecting credit card fraud. As all parameters are verified and well represented in plots. The uniqueness of our approach lies in using the normalization and clustering mechanism of SOM of detecting credit card fraud. This helps in detecting hidden patterns of the transactions which cannot be identified to the other traditional method. With appropriate no of weight neurons and with help of thousands of iterations the network is trained and then result is verified to new transactions. The concept of normalization will help to normalize the values in other fraud cases and SOM will be helpful in detecting anomalies in credit card fraud cases. This unsupervised approach will lead a bank to manage frauds and helps try to avoid the occurrence of fraud as early as possible.

Classes	FPR in SOM	FPR in ID3
Test1	40	42
Test 2	35.71	37
Test 3	29.41	32
Test 4	23.52	28



Classes	TPR in SOM	TPR in ID3
Test	75	60
Test 2	81	65
Test3	88	70
Test 4	92.5	72.5

Figure 3: Results

References

[1] Chen, Q and Aickelin, U. "Dempster Shafer for anomaly detection", Proceedings of the International Conference on Data Mining DMIN 2006, Las Vegas, USA, pp. 232-238(2006).

[2] Kevin J. Leonard. "The development of rule based expert system model for fraud alert in consumer credit", ELSEVIER, 1993, Canada, pp. 350-356 (1993).

[3] Wang, Gang, Hao, Jinxing, Ma, Jian and Huang, "A new approach to intrusion detection using ANN and fuzzy clustering", Original research article expert system with applications, 37(9), pp. 6225-6232(2010).

[4] Eskin E, Arnold, A. Preau, M. Protony, L. Stolfio, "A geometric framework for unsupervised anomaly detection: detecting intrusion in unlabeled data", In Data mining for security application, kluwer(2002).

[5] Zakia, F. and Alkira, M. "Unsupervised outlier detection in time series data" Proceedings of the second International Special workshop on databases for next generation Researchers SWOD2006, pp. 51-56(2006).

[6] Cutrhie, D, Allison and Wilks Y., "Unsupervised Anomaly Detection", IJCAI, pp. 1624-2162(2007).

[7] Dipti Thakur ans Shalini Bhatia, "Distributed Data Mining Approach to Credit Card Fraud Detection", SPIT-IEEE International Conference Mumbai, India, Vol-4, 48(2003).

[8] Shelly Xiaonan, Banzhaf, W. "The use of computational intelligence in intrusion detection System: a review, "Review Article Applied Soft Computing" pp. 1-35(2010).

[9] John T. S Quah, M. Sriganesh, "Real-time Credit card fraud detection using computational intelligence", ELSEVIER, pp. 1721-1732(2008).