ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 7, Issue 7, July 2018

PERFORMANCE EVALUATION FOR CREDIT CARD FRAUD DETECTION USING FREQUENT PATTERN BASED GENETIC ALGORITHM HIERARCHICAL PROCESSING

[1]Vasundhara Punj;

srcvasu2011@gmail.com

Student of computer science and engineering

[2]Prashanta Kumar Bhuyan

Faculty of computer science and engineering

HRIT ,Ghaziabad, Uttar Pradesh

ABSTRACT

As credit card progresses the most popular mode of payment for both online as well as regular purchase, cases of fraud related with it are also increasing. This algorithm is an optimization technique and evolutionary search based on the principles of genetic and natural selection, heuristic used to solve high complexity computational problems. If an incoming credit card transaction is not accepted by the trained HMM with satisfactorily high probability, it is considered to be fraudulent. In this work we also classify the data set with the help of *K-MEAN NEIGHBOR*. At the same time, we try to ensure that genuine transactions are not rejected by using an enhancement to it LUHN & K-Near Neighbor mean with genetic optimization. In further segments we compare different methods for fraud detection and prove that why Genetic optimization more preferred method using MATLAB 2014Ra.

keyword: hmm, luhn, k-mean neighbor, credit card fraud, genetic optimization, frequent pattern etc.

1. INTRODUCTION

1.1 Credit card

Credit Card Fraud is one of the biggest mechanisms of executing a fraud. Credit card fraudsters employ a large number of modus operandi to commit fraud. In simple terms, Credit Card Fraud is defined as:

Credit card frauds are committed in the following ways:

- An act of criminal deception (mislead with intent) by use of unauthorized account and/or personal information
- Illegal or unauthorized use of account for personal gain
- Misrepresentation of account information to obtain goods and/or services.

Contrary to popular belief, merchants are far more at risk from credit card fraud than the cardholders. While consumers may face trouble trying to get a fraudulent charge reversed, merchants lose the cost of the product sold, pay chargeback fees, and fear from the risk of having their merchant account closed.

1.2. Fraud Detection System

Credit card purchase amount will be checked with spending profile of user. By transition probabilistic calculation based on HMM, it concludes whether the transaction is real or fraud. If transaction may be concluded as fraudulent transaction then user must enter security information. This information is related with credit card (like account number, security question and answer which are provided at the time of registration). If transaction will not be fraudulent then it will direct to give permission for transaction. If the detected transaction is fraudulent then the Security information form will arise. The flowchart of proposed module is shown in Figure 1.

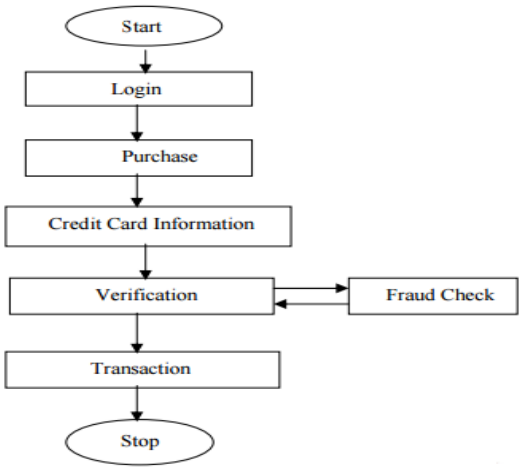

Fig.1: Flowchart of HMM module for credit card fraudulent detection

1.3. K-mean neighbor for fraud detection

K-nearest neighbors is a classification (or regression) algorithm that in order to determine the classification of a point combines the classification of the K nearest points. It is supervised because you are trying to classify a point based on the known classification of other points. K Nearest Neighbor (KNN from now on) is one of those algorithms that are very simple to understand but works incredibly well in practice. Also it is surprisingly versatile and its applications range from vision to proteins to computational geometry to graphs and so on . Most people learn the algorithm and do not use it much which is a pity as a clever use of KNN can make things very simple.

1.4. LUHN algorithm

We are using Luhn Algorithm [59] for card number validation. It was designed to protect against accidental errors, not malicious attacks. Most credit cards and many government identification numbers use the algorithm as a simple method of distinguishing valid numbers from collections of random digits.

2. LITERATURE SURVEY

Saleh Alehalfuraih Richard et al [4] proposed architecture that focuses on the use of Trusted Email mechanism to prevent credit card fraud. It prevents fraudulent transactions of soft-products. It basically consists of a email solution which identifies and authenticates the online customer. It not only prevents fraudulent transactions but also resolves disputes [4].

Abhinav Srivastava et al describe the "Credit card fraud detection method by using Hidden Markov Model (HMM)", [7]. In this Thesis, they model the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of frauds. An HMM is initially trained with the normal behaviour of a cardholder.

S.Ghosh and Douglas L.Reilly et al describes the "Credit card fraud detection With Neural Network", [8].In this Thesis they using data from a credit card issuer, a neural network based fraud detection system was trained on a large sample of labeled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent two-month period of time. The neural network was trained on examples of fraud due to lost cards, stolen cards, application fraud,counterfeit fraud, mail-order fraud and NRI (non-received issue) fraud. The network detected significantly more fraud accounts (an order of magnitude more) with significantly fewer false positives (reduced by a factor of 20) over rule based fraud detection procedures.

Sunil S Mhamane et al describes the "Use of Hidden Markov Model as Internet Banking Fraud Detection", [5]. In this Thesis they explained about how Fraud is detected using Hidden Markov Model also care has been taken to prevent genuine Transaction should not be rejected by making use of one time password which is generated by server and sent to Personal Mobile of Customer.

Pankaj Richhariya describes "A Survey on Financial Fraud Detection Methodologies",[6].The Thesis details as follows. Owing to levitate and rapid escalation of ECommerce, cases of financial fraud allied with it are also intensifying and which results in trouncing of billions of dollars worldwide each year.Credit card fraud detection has drawn a lot of research interest and a number of techniques, with

special emphasis on data mining and neural networks, have been suggested.

Ghosh and Reilly [4] have proposed credit card fraud detection with a neural network. They have built a detection system, which is trained on a large sample of labeled credit card account transactions. These transactions contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, and non-received issue (NRI) fraud.

Recently, Syeda et al. [5] have used parallel granular neural networks (PGNNs) for improving the speed of data mining and knowledge discovery process in credit card fraud detection. A complete system has been implemented for this purpose.

Stolfo et al. [6] suggest a credit card fraud detection system (FDS) using Meta learning techniques to learn models of fraudulent credit card transactions. Metalearning is a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. A Metaclassifier is thus trained on the correlation of the predictions of the base classifiers. The same group has also worked on a cost-based model for fraud and intrusion detection [7]. They use Java agents for Metalearning(JAM), which is a distributed data mining system for credit card fraud detection. A number of important performance metrics like True Positive—False Positive (TP-FP) spread and accuracy have been defined by them.

Aleskerov et al. [8] present CARDWATCH, a database mining system used for credit card fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases.

3. SYSTEM MODEL

During the credit card transaction, the fraudulent transaction is detected and the number of false alert is being minimized by using genetic algorithm. Instead of maximizing the numbers of correctly classified transactions, we defined an aim of function where the miss classification costs are values and thus correct classification of some transactions are more important than correctly identifying the others. The huge amount of losing due to fraudulent transaction and the awareness of the relation between loss and the available limit have to be reduced.

4.1 LUHN ALGORITHM

The formula generates a check digit, which is typically appended to a partial account number to generate the full account number. This account number must pass the following algorithm (and the check digit chosen and placed so that the full account number will)Starting with the second to last digit and moving left, double the value of all the alternating digits.

3.1.1. Algorithm

The algorithm proceeds in three steps. Firstly, every second digit, beginning with the next-to-rightmost and proceeding to the left, is doubled. If that result is greater than nine, its digits are summed (which is equivalent, for any number in the range 10 through 18, of subtracting 9 from it). Thus, a 2 becomes 4 and a 7 becomes 5.

Secondly, all the digits are summed. Finally, the result is divided by 10. If the remainder is zero, the original number is valid. The following is wikicode, a proposed pseudocode for use in many articles.

```
function form Luhn(string purportedCC)
 {
int sum := 0
int nDigits := length(purported CC)
 int parity := nDigits modulus 2
 for I from 0 to nDigits - 1 {
 int digit := integer(purportedCC[I])
 if I modulus 2 = parity
digit := digit × 2
 if digit > 9 digit := digit – 9
 sum := sum + digit
```

}

return (sum modulus 10) = 0

}

4.2 K- NEAREST NEIGHBOR ALGORITHM

The concept of credit card fraud detection by using a data stream outlier detection algorithm which is based on reverse k-nearest neighbors (SODRNN).

The distinct quality of SODRNN algorithm is it needs only one passes of scan. Whereas traditional methods need to scan the database many times, it is not suitable for data stream environment [58]. The performance of KNN algorithm is influenced by three main factors [Mohammed J. Islam]:

• The distance metric used to locate the nearest neighbors.

• The distance rule used to derive a classification from k-nearest neighbor.

• The number of neighbors used to classify the new sample.

4.3 HMM Algorithm

Hidden Markov model is one of the fraud detection techniques. This technique uses a finite set of states to verify the transactions.

Each state has its own probabilistic distribution that is been used to verify the given transaction. The outcome of the state is given to the observer to determine its output. The internal states are hidden from the observer, Hence it is called as Hidden Markov model (HMM). The different states in the Hidden Markov Model can be represented as shown in following diagram

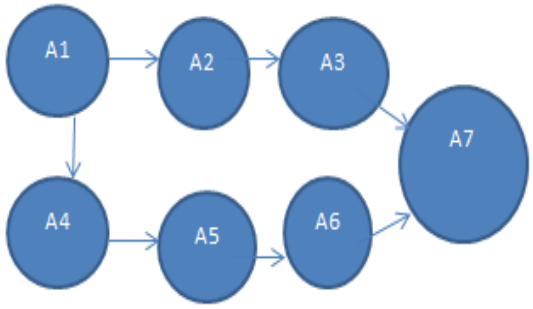

Fig 4.3: State representation in HMM

Hidden Markov Model can be defined with the help of following elements

a. N, the number of positions,

b. M, the number of observation symbols,

c. A, matrix of state transition probabilities,

d. B, matrix of observation emission probability distribution,

e. matrix of prior probabilities

4.4 Decision Tree

A decision tree consists of nodes and branches. The starting node is usually referred to as the root node. Each node is labeled with a feature name and each branch leading out of it is labeled with one or more possible values for that feature. This algorithm uses the concept of information entropy to determine the best node for the tree to branch to. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision. Entropy for a set of examples, S, for one variable can be calculated as follows:

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Where i is the outcome state, pi is the probability of outcome state i, and c is the number of outcome states.

Entropy for two variables can be calculated as follows:

$$E(S,A) = \sum_{v \in A}^{c} \frac{|S_v|}{|S|} E(S_v)$$

Where v is the state of the second variable, A is the set of examples of the second variable, Sv is the size of the subset in state v, and S is the size of the entire set.

Finally the information gain is defined as:

$$Gain(S,A) = E(S) - \sum_{v \in A}^{c} \frac{|S_v|}{|S|} E(S_v)$$

The entropy of an attribute represents the expected amount of information that would be needed to specify the classification of a new instance. Therefore the attribute with the largest amount of information gained would be selected as the splitting attribute. The decision tree is stopped whenthe data cannot be split any further. Ideally, the process is repeated until all leaf nodes are pure, that is, when they contain instances that have the same classification.

4.5. GAHP-FP (Frequent pattern and genetic algorithm hierarchical pattern)

Fraud pattern that's represented previously and confirmed as fraud transactions that's facilitate studding fraudster's behavior. An enhancement for the proposed algorithm of Fraud Miner has been proposed. Using this algorithm provide more chance for easily fraud detection as the fraudsters always behaving same as customer behaviors instead of study fraudster behavior the customer frequent behavior will be identified from his legal or previously confirmed transactions being fraud. A performance comparison with other algorithms has been carried out.

The Experiment process has four steps.

Step1. Input group of data credit card transactions, every transaction record with n attributes, and standardize the data, get the sample finally, which includes the confidential information about the card holder, store in the data set.

Step2. Compute the critical values, Calculate the CC usage frequency count, CC usage location, CC overdraft, current bank balance, average daily spending

Step3. Generate critical values found after limited number of generations. Critical Fraud Detected, Monitorable Fraud Detected, Ordinary Fraud Detected etc. using Genetic algorithm

Step4. Generate fraud transactions using this algorithm. This is to analyze the feasibility of credit card fraud detection based on technique, applies detection mining based on critical values into credit card fraud detection and proposes this detection procedures and its process.

4.5.1. Genetic algorithm

The current values of these parameters have been determined, and critical values are compared with the data set parameters, also maximizes the number of true alerts given that the number of alerts does not exceed a certain level.

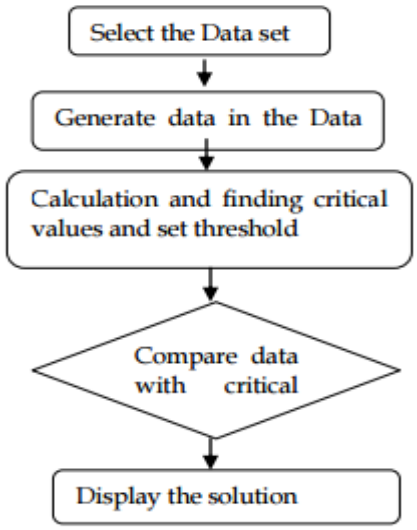

Fig. 4.5.1 The simple method of Genetic Algorithm

Genetic algorithm is the procedure is repeated until a pre specified number of generations has passed, and the best solution found. It is parametric procedure and it needs to be problem undertaken to get a better performance. To calculate the CC usage frequency count, CC usage location, CC overdraft, current bank

balance, average daily spending etc. as shown in Fig.4.5.2

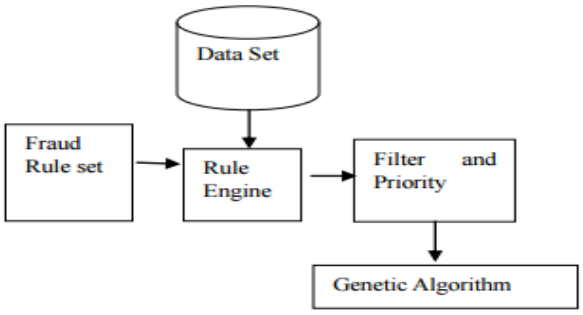

Fig. 4.5.2 System Design

As shown in Fig 5.3, Genetic algorithms are evolutionary algorithms which aim at obtaining better solutions as time progresses. It also been used in data mining mainly for variable selection and are mostly coupled with other data mining algorithms. In this study, we are solving the classification problem by using only a genetic algorithm solution.

The initial population is selected randomly from the sample space which has many populations. The fitness value is calculated in each population and is sorted out. In selection process is selected through tournament method. The Crossover is calculated using single point probability. Mutation mutates the new offspring using uniform probability measure. In elitism selection the best solution are passed to the further generation. The new population is generated and undergoes the same process it maximum number of generation is reached .

Pseudo code of genetic algorithm

- ✓ ***INITIALIZE THE POPULATION***
- ✓ ***EVALUATE INITIAL POPULATION***
- ✓ ***REPEAT***
- ✓ ***PERFORM COMPETITIVE SELECTION***
- ✓ ***APPLY GENETIC OPERATORS TO GENERATE NEW SOLUTIONS***
- ✓ ***EVALUATE SOLUTIONS IN THE POPULATION***
- ✓ ***UNTIL SOME CONVERGENCE CRITERIA IS SATISFIED.***

5. RESULT& IMPLEMENTATION

For the purpose of fraud detection, HMM to describe the behavior of user are constructed. First, a k-mean neighbor is constructed to model behavior under the assumption that the user is fraudulent (F) and another model under the assumption the user is a legitimate (NF). The 'fraud net' is set up by using expert knowledge. The 'user net' is set up by using data from non-fraudulent users. During operation user net is adapted to a specific user based on emerging data. By inserting evidence in these networks and propagating it through the network, the probability of the measurement x less than two above mentioned hypotheses is obtained. This means, it gives judgments to what degree observed user behavior meets typical fraudulent or non-fraudulent behavior. These quantities we call $p(x \mid NF)$ and $p(x \mid F)$. By postulating the probability of fraud P (F) and P (NF) = 1-P(F) in general and by applying Bayes' rule, it gives the probability of fraud, given the measurement x,

$$P(F \mid x) = \frac{P(F)\, p(x \mid F)}{p(x)}$$

Where the denominator p(x) can be calculated as

$$P(x) = P(F)p(x \mid F) + P(NF)p(x \mid NF)$$

The fraud probability $P(F \mid x)$ given the observed user behavior x can be used as an alarm level. On the one hand, Bayesian networks allow the integration of expert knowledge, which we used to initially set up the models [9]. On the other hand, the user model is retrained in an unsupervised way using data. Thus our Bayesian approach incorporates both, expert knowledge and learning.

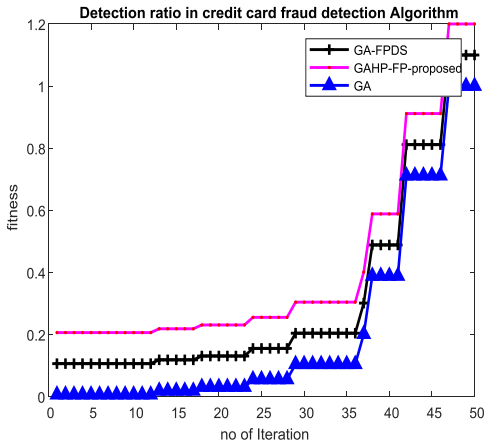

Figure 5.1: Detection ratio in credit card fraud detection algorithm

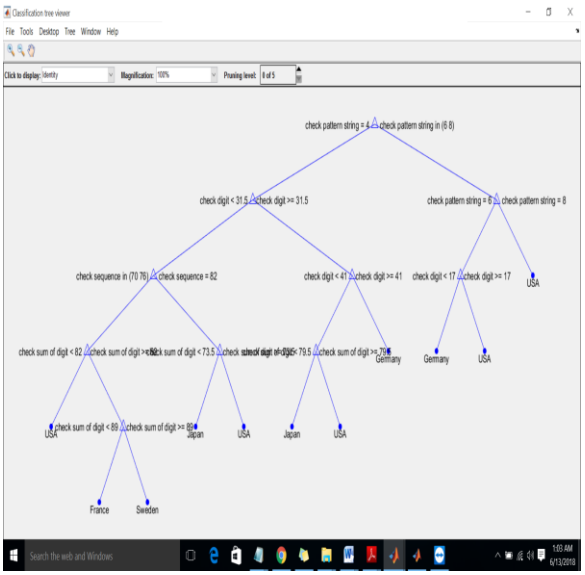

Figure 5.2: Classification review of different country

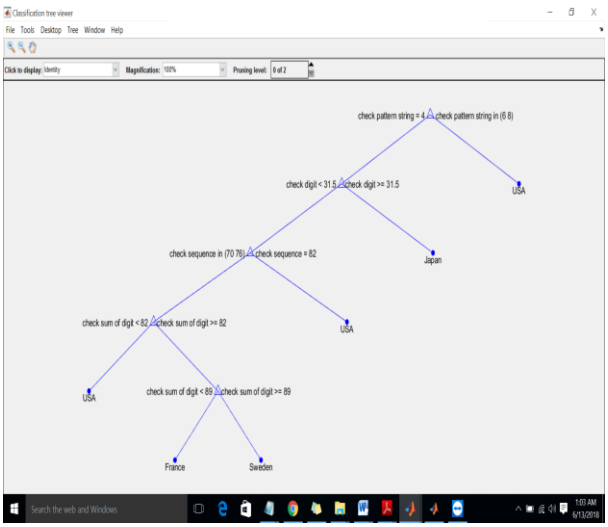

Figure 5.3: Classification review

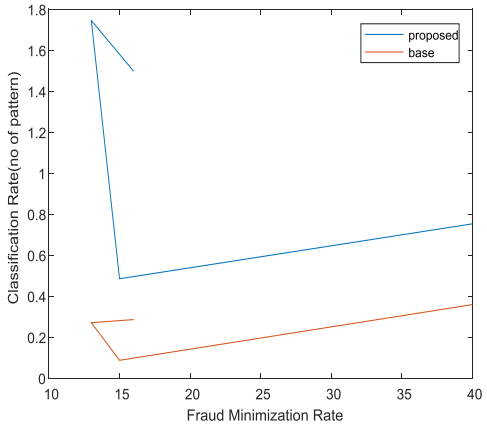

Figure 5.4: Comparison between Base and proposed of classification rate (no. of pattern)

6. CONCLUSION

we present to find the detection of credit card fraud mechanism and examine the result based on the principles of this algorithm. The genetic algorithm that are being used to execute credit card fraud how credit card fraud impact on financial institution as well as merchant and customer, fraud detection technique by genetic algorithm. The Genetic algorithms are evolutionary algorithms in which the aim is to obtain the better and optimal solutions. In this study fraud detected and fraud transactions are generated with the given sample data set. If this

algorithm is applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted soon after credit card transactions by the banks. And a series of anti-fraud strategies can be adopted to prevent banks from great losses before and reduce risks.

FUTURE WORK

Neural Network is train on this data but the problem arises at the initial stages when very few or not at all initial transaction has been made, how will we train KNN when only few or no data is available to train the network because in order to make a Neural Network to predict we must have some pattern through which NN can get train and make prediction.

REFERENCES

[1]Credit Card Security and E-payment, Enquiry into credit card fraud in E-Payment, Jithendra Dara, Luleå University of Technology, 2006 .

[2]Unawed by fraud: new techniques and technologies have been enlisted in the fight against online fraud, by Micci-Barreca, Daniele, Security Management, Electronic Commerce, Sept, 2003.

[3] S Alfuraih, "Location of Trusted Email for Detection of Credit Card Fraud in SoftProducts E-Commerce", 2004.

[4] Saleh I. Alfuraih, Nie n T. Sui and Dennis McLeod, "Using Trusted Email to Detect Credit Card Frauds in Multimedia Products", 2002

[5] Sunil S Mhamane and L.M.R.J Lobo "Use of Hidden MarkovModel as Internet Banking Fraud Detection" International Journalof Computer Applications (0975 – 8887) Volume 45– No.21, May 2012

[6] Pankaj Richhariya et al "A Survey on Financial Fraud DetectionMethodologies" BITS,Bhopal," International Journal of ComputerApplications (0975 – 8887) Volume 45 No.22, May 2012.

[7] Srivastava, Abhinav, Kundu, Amlan, Sural, Shamik and Majumdar,Arun K., (2008) "Credit Card Fraud Detection Using HiddenMarkov Model", IEEE Transactions on Dependable and SecureComputing, Vol. 5, No. 1, pp. 37-48..

[8] S. Ghosh and D.L. Reilly, "Credit Card Fraud Detection with aNeural-Network," Proc. 27th Hawaii Int'l Conf. SystemSciences:Information Systems: Decision Support and Knowledge-Based Systems,vol. 3, pp. 621-630, 1994.

.

[9] S. Ghosh and D.L. Reilly, "Credit Card Fraud Detection with aNeural-Network," Proc. 27th Hawaii Int'l Conf. System Sciences:Information Systems: Decision Support and Knowledge-Based Systems,vol. 3, pp. 621-630, 1994.

[10] M. Syeda, Y.Q. Zhang, and Y. Pan, "Parallel Granular Networksfor Fast Credit Card Fraud Detection," Proc. IEEE Int'l Conf. FuzzySystems, pp. 572-577, 2002.

[11] S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan,"Credit Card Fraud Detection Using Meta-Learning: Issues andInitial Results," Proc. AAAI Workshop AI Methods in Fraud and RiskManagement, pp. 83-90, 1997.

[12] S.J. Stolfo, D.W. Fan, W. Lee, A. Prodromidis, and P.K. Chan,"Cost-Based Modeling for Fraud and Intrusion Detection: Resultsfrom the JAM Project," Proc. DARPA Information Survivability Conf.and Exposition, vol. 2, pp. 130-144, 2000.

[13] E. Aleskerov, B. Freisleben, and B. Rao, "CARDWATCH: ANeural Network Based Database Mining System for Credit CardFraud Detection," Proc. IEEE/IAFE: Computational Intelligence forFinancial Eng., pp. 220-226, 1997.

[14] M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud DensityMap for Effective Credit Card Fraud Detection," Proc. Int'l Conf.Intelligent Data Eng. and Automated Learning, pp. 378-383, 2002.

[15] W. Fan, A.L. Prodromidis, and S.J. Stolfo, "Distributed DataMining in Credit Card Fraud Detection," IEEE Intelligent Systems,vol. 14, no. 6, pp. 67-74, 1999.

[16] R. Brause, T. Langsdorf, and M. Hepp, "Neural Data Mining forCredit Card Fraud Detection," Proc. IEEE Int'l Conf. Tools withArtificial Intelligence, pp. 103-106, 1999.