ISSN: 2229-6956(ONLINE)
DOI: 10.21917/ijsc.2012.0061

BEHAVIOR BASED CREDIT CARD FRAUD DETECTION USING SUPPORT VECTOR MACHINES

V. Dheepa[1] and R. Dhanapal[2]

[1]*Research and Development Centre, Bharathiar University, India*
E-mail: dsvdeepasaro@gmail.com
[2]*Department of Computer Applications, Easwari Engineering College, India*
E-mail: drdhanapal@gmail.com

Abstract

Along with the great increase of internet and e-commerce, the use of credit card is an unavoidable one. Due to the increase of credit card usage, the frauds associated with this have also increased. There are a lot of approaches used to detect the frauds. In this paper, behavior based classification approach using Support Vector Machines are employed and efficient feature extraction method also adopted. If any discrepancies occur in the behaviors transaction pattern then it is predicted as suspicious and taken for further consideration to find the frauds. Generally credit card fraud detection problem suffers from a large amount of data, which is rectified by the proposed method. Achieving finest accuracy, high fraud catching rate and low false alarms are the main tasks of this approach.

Keywords:

Data Mining, Classification, Fraud Detection, Support Vector Machine, E-commerce

1. INTRODUCTION

Payments using credit cards have increased in recent years. It may be used in online or in regular shopping. Now-a-days credit card payments are necessary and convenient to use. Due to the increase of fraudulent transactions, there is a need to find the efficient fraud detection model.

Fraud is increasing dramatically with the expansion of modern technology [1].The amount of information in the modern world is also explosive [2]. The promising way to detect the fraud is to analyze the spending behavior of the cardholder. Detecting the fraud means identifying the suspicious one [3]. If any abnormality arises in the spending behavior then it is considered as suspicious and taken for further consideration. In this Paper, a behavior based approach using support vector machines is applied for fraud detection.

Support Vector Machine (SVM) is an active research area and successfully solves classification problems in noisy and complex domains. SVM played a major role in the area of machine learning due to its excellent generalization performance in a wide range of learning problems, such as hand written digit recognition, classification of web pages and face detection. The Problem of over fitting is very less in SVM applications. The problems of multi-local minima and curse of dimensionality rarely occurs in SVM. Support Vector machines originated from statistical learning theory, a theoretical base of statistical inference. Recently SVM has been employed in business applications such as Credit rating analysis [4], bankruptcy prediction [5] and time series prediction and classification [6]. In this work, SVM is applied for fraud detection to classify and predict the data.

2. RELATED WORK

There was a lot of research work carried out for credit card fraud detection. CARDWATCH, a data base mining system proposed by Aleskerov et al. [7]. It was based on neural networks. In this model, customers past transactions are trained in the neural network. Then the network checks the current spending pattern with the past data, if deviations appear then it is considered as suspicious. Zhang Yongbin et al. [8] suggested a behavior based credit card fraud detection model. Here they use the historical behavior pattern of the customer to detect the fraud. The transaction record of a single credit card is used to build the model. In this model, unsupervised Self organizing map method is used to detect the outliers from the normal ones.

Chuang et al. [9] developed a model based on data mining. They used the web services to exchange data between banks and fraud pattern mining algorithm for detection. With the proposed scheme participant banks can share the knowledge about fraud patterns in a heterogeneous and distributed environment and further enhance their fraud detection capability and reduce financial loss. Wen-Fang Yu et al. [10] proposed an outlier mining method to detect the credit card frauds. Definitions of Distance based outliers are referred and the outlier mining algorithm was created. This model detects outlier sets by computing distance and setting threshold of outliers. It efficiently detects the overdrafts and is also used to predict the fraudulent transactions.

Tao Guo et al. [11] applied the neural data mining method. This model is based on customer’s behavior pattern. Deviation from the usual behavior pattern is taken as an important task to create this model. The neural network is trained with the data and the confidence value is calculated. The credit card transaction with low confidence value is not accepted by the trained neural network and it is considered as fraudulent. If the confidence value is abnormal, then again it is checked for additional confirmation. The detection performance is based on the setting of threshold.

Suvasini Panigrahi et al. [12] suggested a fusion approach. It consists of four components namely, rule based filter, Dempster-Shafer Adder, transaction history database and Bayesian learner. Rule based filter is used to find the suspicion level of the transaction. Dempster-Shafer Theory is used to compute the initial belief which is based on the evidences given by the rule based filter. The transactions are classified as normal, abnormal or suspicious depending on this initial belief. Once a transaction is found to be suspicious, belief is further strengthened or weakened according to its similarity with fraudulent or genuine transactions history using Bayesian learning.

Abhinav Srivastava et al. [13] developed the hidden Markov model (HMM) to detect the credit card fraud. An HMM is initially trained with the normal behavior of the cardholder. If the current transaction is not accepted by the trained HMM with high probability, it is considered to be fraudulent. Vladimir et al. [14] applied self organizing map algorithm to create a fraud detection model. The pattern of legal and fraudulent transactions is observed from the earlier transactions and it is created based on the neural network training. If a new transaction does not match to the pattern of legal cardholder or is similar to the fraudulent pattern it is classified as suspicious for fraud. Chen et al. [15] proposed the online questionnaire method to collect the transaction data of the users. A SVM is trained with the data and the questionnaire responded transaction is used to predict the new transactions.

3. FRAUD DETECTION

3.1 FRAUD DETECTION PROBLEM

Fraud is defined as criminal deception. The purpose of fraud may be to obtain goods without paying or to obtain unauthorized funds from an account. Fraud can either be prevented or detected. In prevention, precaution activities are made to reduce the fraud. If the fraud prevention fails the problem of detection is taken for consideration. Fraud detection is identifying wrongful, suspect and illegitimate behavior. There are a lot of procedures used for fraud detection. The main target is defending the transactions from illegal use and maximizes the correct predictions.

The credit card fraud can be of two types: offline fraud and online fraud. The fraud begins either with the theft of the credit card or the compromise of data associated with the account, including the card account number or other information that would routinely and necessarily be available to a merchant during a legitimate transaction. The offline fraud associated with the theft of the credit card. The physical card is stolen by the unauthorized person. Then purchases made by him by using the stolen card. He may use it to purchase until the usage of card is cancelled. If the user does not know about the theft, then it is a great loss to him and to the corresponding financial institutions [16]. In online fraud, the physical card is not needed only the information about the card is enough to purchase. Thus here the fraudster simply needs some important card details. Stealing the information from the user is called Identity Theft. In this type fraud the transactions are done through phone or internet. Generally, the genuine cardholder is not responsive if someone else has seen or stolen his card information.

3.2 CHALLENGING ISSUES IN FRAUD DETECTION

The datasets are extreme imbalance and highly skewed. The genuine transactions dominate than fraudulent transactions. The fraudulent events occur rarely. So it is difficult to find the fraudulent. If the fraudulent transaction is consider as legal then it will cause great loss.

The huge amount of datasets and the dimensionality is very high. It is not an easy process to handle the massive amount of data efficiently. The scalable machine learning system is needed to process the large amount of data.

The real data is not shared for the number of reasons such as to maintain the privacy of the user.

Generally the misclassification cost is high for these detections. Efficient measure should take to reduce the misclassification cost.

3.3 BEHAVIOR BASED MODEL

The key concept in fraud detection is to analyze the spending behaviors of the user. If any discrepancies occur with the respect to the usual spending behavior, then it is considered as suspicious behavior. And it is taken for further consideration. The behavior of spending varies from person to person. Fraud detection based on the analysis of existing spending behavior of cardholder is a promising way to find the credit card frauds.

Behavior based fraud detection model means that the data use in the model are from the transactional behavior of cardholder directly or derived from them. Each person may have a different spending behavior pattern. Most of the existing fraud detection methods use the behavior pattern as measure to find the destruction in the transactions. Based on the spending pattern the customer’s usual activities such as transaction amount, billing address etc are learned. Some of the count measures to suspect the behaviors are the variation of billing address and shipping address, maximum amount of purchase, large transaction done far away from the living place etc. Like that the behaviors deviate from the normal ones are suspected and taken for further consideration.

4. OVERVIEW OF SUPPORT VECTOR MACHINE

Support vector machine is a method used in pattern recognition and classification. It is a classifier to predict or classify patterns into two categories; fraudulent or non fraudulent. It is well suited for binary classifications. As any artificial intelligence tool, it has to be trained to obtain a learned model. SVM has been used in many classification pattern recognition problems such as text categorization [17], bioinformatics [18] and face detection [19]. SVM is correlated to and having the basics of non-parametric applied statistics, neural networks and machine learning. SVM is described in the following passage [20], [21].

It was developed from the theory of Structural Risk Minimization. In a binary classification problem, the decision function of SVM is shown in Eq.(1),

$$f(x) = \mathrm{sgn}(x.w) + b \qquad (1)$$

where, x is the input vector which contains weight and b is a constant. Eq.(1) is used to find the decision boundary between two classes. The parameter values of w and b have to be learned by the SVM on the training Phase and b are derived by maximizing the margin of separation between the two classes. The criterion used by SVM is based on the margin maximization between the two classes.

The margin is the distance between the two hyper planes. To find the hyper plane H:y = *w.x* + *b* = 0 and two hyper planes H1:*y* = *w.x* + *b* = +1 and H2:*y* = *w.x* + *b* = -1.The threshold separating the two classes is H and the two margin boundaries are H1 and H2. Then the margin is 2 / ||*w*|| , where ||*w*|| is the norm of the vector *w*. In non perfectly separable case the margin is soft. That there is a chance of misclassification error. The misclassification errors should be minimized. It is minimized by introducing the slack variable ξ_i .If ξ_I = 0 then the classes are correctly classified. Let ξ_i is non-negative slack variable for misclassifications.

Let *y* is the indicator of the class, where in the case of fraud detection *y* = 1 for the positive class and *y* = -1 is the class for the negative class.

SVM requires that either *x.w* + *b* >= 1 – ξ_i or *x.w* + *b* >= -1 + ξ_i which can be summarized with Eq.(2),

$$y_i\left(w.x+b\right)>=1-\xi i \tag{2}$$

where, *i* = 1,2,…, *n*

The optimization problem for the calculation of *w* and *b* can thus be defined by the Eq.(3),

$$Min\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\xi_i \tag{3}$$

Subject to $y_i(w.x + b) >= 1 - \xi_i,\qquad \xi_i >= 0$

By minimizing the $\|w\|^2/2$ the complexity of SVM is reduced and by minimizing the slack variable the misclassification errors are reduced. *C* is a regularization parameter which weighs the classification errors. And it is the tradeoff between the two classes. The constrained optimization problem is solved by using the Lagrange function 4,

$$L(w,b,\xi,\alpha,\beta)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\alpha_i\{y[wx+b]-1+\xi_i\} \\ -\sum_{i=1}^{n}\beta_i\xi_i \tag{4}$$

The solution of this optimization problem is obtained by minimizing *w*, *b* & ξ and maximizing α &β. It is better to solve the problem by introducing the dual formulation in Eq.(5),

$$\max_{\alpha\beta} w(\alpha,\beta)=\max_{\alpha\beta}\left\{\min_{w,b,\xi}\left(w,b,\xi,\alpha,\beta\right)\right\} \tag{5}$$

By substituting this, the problem is transformed into its dual formulation, specified by,

$$\max\left\{\sum_{\mathrm{i=1}}^{\mathrm{n}}\alpha-\sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{j=1}}^{\mathrm{n}}\alpha_{\mathrm{i}}\alpha_j y_i y_j\left\langle x_i,y_j\right\rangle\right\} \tag{6}$$

and is maximized under the constraints,

$$\sum_{i=1}^{n}\alpha_i y_i=0 and 0\le\alpha_i\le C \text{ for } i = 1,2,..,n.$$

The Kuhn-Tucker condition in Eq.(7) is applied to Eq.(6),

$$\alpha_i\{y_i[w.x_i + b] - 1 + \xi\} = 0_i \tag{7}$$

where, *i* = 1,2,..., *n*.

The Lagrange vectors are the support vectors needed to describe the hyper plane. In linearly separable data, all support vectors all lay on the margin. The decision boundary is determined by the Eq.(8),

$$f(x)=\sum_{i=1}^{Ns}\alpha_i y_i\left\langle x,x_i\right\rangle+b \tag{8}$$

where, *x* is the input vector, (*x*, *x*$_i$) is the inner product, *Ns* is the number of support vectors, and *b* is the bias term.

If the data are non linear, then kernel trick is used to map the input vector to higher dimensional feature space. Kernel functions can be used in many applications as they provide a simple bridge from linearity to non-linearity for algorithms which can be expressed in terms of dot products. Any function that satisfies Mercer's condition by vapnik is used as a kernel function.

The kernel function is introduced as follows [22],

$$\left\langle x_i,x_j\right\rangle\to k(x_i,x_j) \tag{9}$$

By using the kernel function, the input vectors are mapped to a higher dimensional feature space. Some of the kernel functions used in SVM is,

Linear Kernel Function:

$$k(x_i,x_j) = x_i.x_j \tag{10}$$

The Linear kernel is the simplest kernel function. It is given by the inner product <*x*, *y*> plus an optional constant *c*.

Sigmoid Kernel Function:

$$k\ (x_i,x_j) = \tanh(\gamma x_i x_j + r) \tag{11}$$

γ and r are the two parameters are used in sigmoid kernel. If γ > 0, then *γ is acting* as a scaling parameter of the input data, and r as a shifting parameter that controls the threshold of mapping. If γ < 0, the dot-product of the input data is not only scaled but reversed.

Polynomial Kernel Function:

$$k(x_i,x_j) = [(x_i.x_j)+1]d \tag{12}$$

The Polynomial kernel is a non-stationary kernel. Polynomial kernels are well suited for problems with normalized training data.

RBF Kernel Function:

$$k(x_i,x_j)=\exp(-\gamma\left\|x_i-x_j\right\|^2) \tag{13}$$

The adjustable parameter *γ* plays a major role in the performance of the kernel, and should be carefully tuned to the problem at hand. If overestimated, the exponential will behave almost linearly and the higher-dimensional projection will start to lose its non-linear power. On the other hand, if underestimated, the function will lack regularization and the decision boundary will be highly sensitive to noise in training data.

In the proposed work, the RBF kernel function is used. This function is more flexible and majority of the SVM applications used the RBF kernel function. The RBF have a parameter known as Gaussian width γ, it controls the width of the RBF kernel function. The performance of SVM can be based on the selection of *C, γ*. These two parameters are most important for the accuracy of classification. If the value of *C* is too small then insufficiency occurs in fitting the training data. If the value of *C* is too large then the over fitting occurs in the training data. Effectively setting the proper values of *C, γ* avoid over fitting and also give good accuracy in classification. The solution of SVM is unique. The non linear data are also easily separated by the use of kernel function. If the parameter values are properly chosen then it provides an effective result. It does not depend on the dimensionality of the input space. If more attention is given to selecting the feature of SVM, it will yield an accurate classification.

5. ROLE OF DATA MINING IN FRAUD DETECTION

Commonly, fraud detection problem is considered as a data mining problem. Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques.

The detection of credit card fraud is generally adopted a classification model. Fig.1 shows the out line of the model. Nowadays data mining is largely applied to the classification and regression problems. The classification is done effectively by the data mining methods. Classification is the process of finding a set of models to differentiate the data concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown.

Supervised learning techniques have been the dominated methods used for fraud detection. Supervised learning models are produced by using the knowledge of already classified data and inductively finding a predictive pattern. Otherwise the supervised learning is to build a concise model of the distribution of the class label in terms of the predictor features.

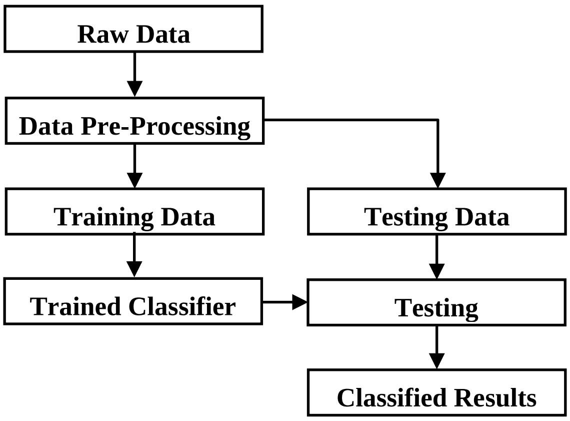

Fig.1. Outline of the Model

The resulting classifier is then used to assign class labels to the testing instances where the values of the predictor features are known but the value of the class label is unknown. A classifier is created from the learned behavior pattern and trained. The test sample is applied to the trained classifier. Using data mining in fraud detection gives various benefits. Basically it is used to analyze the complete datasets for fraud. Scalability is possible in data mining. It is used to find the cost effective solution.

6. METHODOLOGY

Analyzing the spending behavior pattern of the customer is a promising way to detect the credit card frauds. If any new transactions are deviate from this behavior, then that is considering as fraudulent transaction. So, in this paper the feature which shows the behavior of the customer is selected from the data set. The behavior character shows the spending patterns are transaction amount, date, time, place, frequency of purchase and billing Address. Based on these details, the model is trained.

In this model SVM is used for classification. It gives good result when small numbers of features are used. Here, only the behavior features are selected for training. The other challenging issue in this problem is handling the massive amount of data. To recover this problem effective feature extraction method is used, which is used for data reduction. Then the data is trained in SVM classifier. The classification performance of SVM is affected by its parameters [23]. By choosing proper values for the parameters *C* and *γ* avoids over fitting and yields a perfect accuracy. In this model, the parameter values are selected by cross validation using Grid search method. Fig.2. show the workflow of the training model.

7. EXPERIMENTS

7.1 DATA PRE PROCESSING AND SELECTION

Firstly the features used in the dataset are converting into numerical data. Feature selection is a very important stage in fraud detection. The features in the data efficiently portray the usage of behavior of an individual credit card account. In this model, the features which interpret the behavior of the customer only are selected for detection. Adding irreverent features make the classifier inefficient. The important features used in this model are shown in Table.1.

Transaction amount is the most important behavior it varies from person to person. Frequency of card usage is calculated from the Date and Time Attributes. Average amount of transactions are calculated from each transactions.

Table.1. Selected Important Features

Feature No.	Feature Name
1	Transaction Amount
2	Date
3	Time
4	Frequency of card usage
5	Place
6	Customer ID
7	Average amount of transactions per month

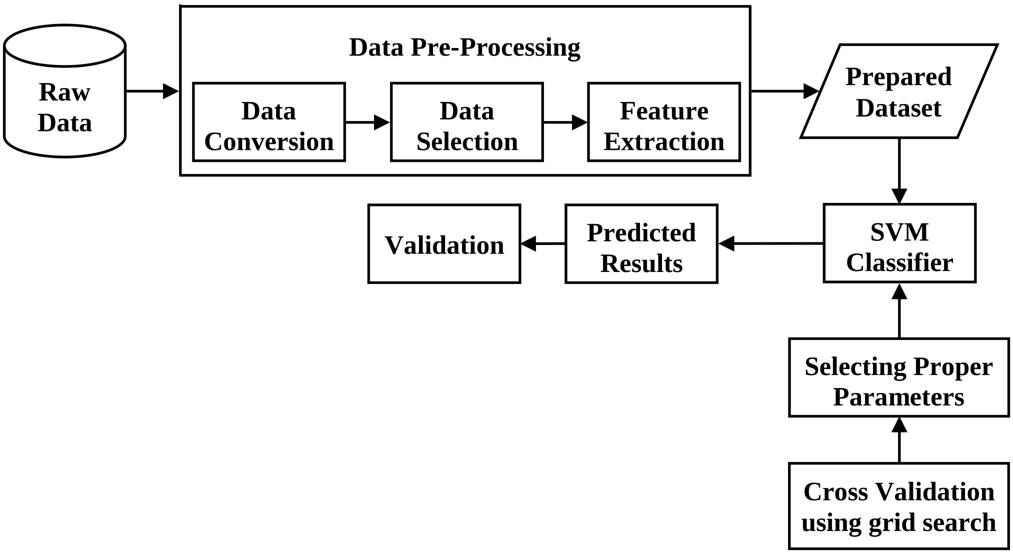

Fig.2. Workflow of the Training Model

7.2 FEATURE EXTRACTION

Feature extraction is special form dimensionality reduction. Here the input data is transformed into a reduced representation set of features. The features represent the relevant characteristics of the input data. Instead of using full size input one may use this reduced representation set. If it is properly chosen then it will give successful task.

Principal component analysis (PCA) is a suitable tool for feature compression. The original feature space is reduced to low dimensional spaces but it will not affect the solution. The computational cost also less for training and testing the SVM because of using PCA.

PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. PCA can be done by Eigen value decomposition of a data covariance matrix, usually after mean centering the data for each feature. The covariance matrix [24] is calculated by the following Eq.(14),

$$Covariance(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - x)(y_i - y) \qquad (14)$$

where, x_i and y_i are the values of variables, x and y are the mean variables, n is number of objects. The results of PCA come in the form of component scores. By using this formulation the principal components are calculated and the training dataset is reduced. This method used to reduce the training time. The adoption of this method makes the classifier to perform in an efficient time manner.

7.3 SVM TRAINING AND CLASSIFICATION

LIBSVM classifier is used for training and classification. Libsvm has lot of functions. The 591 samples selected including 576 positive samples and 15 negative samples. Usually SVM suffers from large number of features. To overcome this problem only selected features are used in this model. If the numbers of features are less and the instances are high then one may have to use the kernel function. In this model RBF kernel function is used. The accuracy is obtained by optimizing the RBF kernel parameter γ and the penalty parameter C.

In this paper, cross validation using grid search method is used to obtain the best γ and C. The V-fold cross validation is used in this process. The training data set is divided into five equal subsets. The classifier is trained five times, not including a single subset at every time.

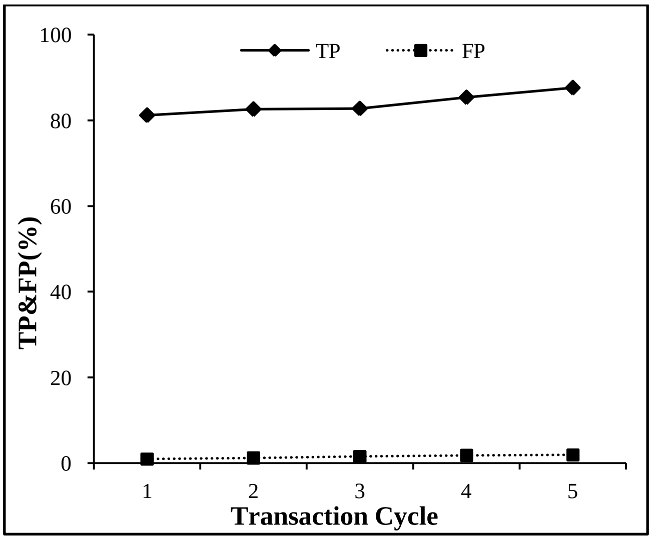

Fig.3. Transaction Rate

The resultant classifier is tested on the excluded subset. The over fitting problem can be recovered by this cross validation [25]. Using grid search method in this one will give better result. In the grid search method, the exponentially growing sequence of C and γ such as $C = 2^{-5}, 2^{-3}, 2^{-15}$ and $\gamma = 2^{-15}, 2^{-13}, 2^{3}$ is used to gain the best cross validation. The obtained parameter values are $C = 2.0$ and $\gamma = 0.81$ gives the highest accuracy.

7.4 RESULTS AND EVALUATION

Even though the accuracy is important, the fraud catching rate and false alarm rate are the better metrics for the fraud detection domain [26]. In this work, the confusion matrix is used for evaluating the fraud catching rate and false alarm rate. The standard confusion matrix format is shown in the following Table.2.

Table.2. Confusion Matrix

		Predicted	
		Positive	**Negative**
Actual	Positive	TP	FN
	Negative	FP	TN

In Confusion Matrix, the column signifies the predicted class and the row signifies the actual class. TP is True Positive (Fraud catching rate) which shows the number of genuine transactions correctly identified as non fraudulent. FP is False Positive (False alarm rate) which gives the number of genuine transactions incorrectly identified as fraudulent. FN is False Negative mistakenly consider fraudulent transaction as genuine. TN is True Negative which shows the number of fraudulent transactions correctly identified as fraudulent. Achieving highest fraud catching rate and lowest false alarm rate is the important task of this model.

The True Positive rate (TP) and False Positive rate (FP) are found by the following Eqs.(15) and (16),

$$TP_{rate} = \frac{TP}{TP + FN} \tag{15}$$

$$FP_{rate} = \frac{FP}{TN + FP} \tag{16}$$

TP_{rate} represents the ratio of positive class that was correctly identified. FP_{rate} represents the ratio of the negative cases that was incorrectly identified as positive. The results of TP_{rate} and FP_{rate} obtained in the proposed approach are shown in Fig.3.

Accuracy represents the ratio of the total number of transactions that were correctly identified. The accuracy of the classifier is calculated by the Eq.(17) and the error rate is calculated by the Eq.(18),

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{17}$$

$$Error\ Rate = \frac{FP + FN}{TP + FP + FN + TN} \tag{18}$$

The accuracy and the error rate of the proposed work are shown in Fig.4.

In the Result, the TP value raises and FP value becomes low. The highest fraud catching rate and low false alarm rates are obtained by selecting the appropriate parameter values. The parameter values are found out by checking the behavior profiles of the cardholders. The parameter values of the proposed work are based on the average amount of transactions and the frequency of the card usage. TP and FP varies for different parameter values are also shown in the Fig.3. This model achieves the accuracy more than 80 percent. Achieving high accuracy is a vital one and reducing the false alarms are also the important tasks in the credit card fraud detection. Too many false alarms restricted the customer from the use of credit card. In the proposed approach false alarm rates are reduced. And also obtaining low false alarm rates will not restrict the fraud catching rate. The proposed model efficiently finds out most of the correct transactions and is well suited for fraud detection.

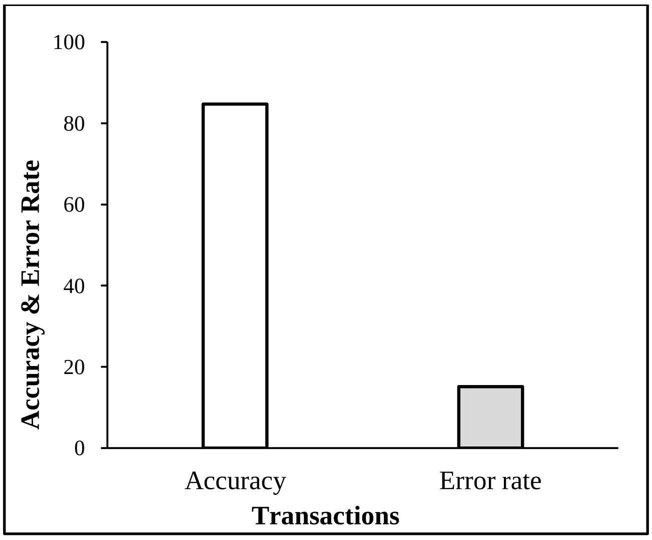

Fig.4. Accuracy and Error rate

8. CONCLUSION AND FUTURE WORK

In this paper, behavior based classification approach using Support vector machine is applied. The proposed method using SVM gives effective performance in fraud detection. Generally SVM deliver a unique solution. By using the kernel, SVM gains flexibility in the form of threshold for separating the data's. Such qualities make the SVM to carry out the classification problem in this complex domain and also it yields a good result. The proposed method gives higher accuracy of detection and is also scalable for handling large volumes of transactions. In future, the cost based support vector machine with effective kernel function will be used to find the fraud detection with lower error rates.

REFERENCES

[1] Tareq Allan and Justin Zhan, "Towards Fraud Detection Methodologies", *IEEE Proceedings of the Fifth International Conference on Future Technology (Future Tech)*, 2010.

[2] R. Dhanapal, "An intelligent information retrieval agent", *ELSEVIER, Knowledge-Based Systems 21*, pp. 466-470, 2008.

[3] V. Dheepa, R. Dhanapal and D. Remigious, "A Novel Approach to Credit Card Fraud Detection Model", *Journal of Computing*, Vol. 2, No. 12, pp. 96, 2010.

[4] Zan Huang, et al., "Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study", *Elsevier-Decision Support Systems*, Vol. 37, pp. 543-558, 2004.

[5] Kyung-Shik Shin, Taik Soo Lee and Hyun-jung Kim, "An application of support vector machines in bankruptcy prediction model", *ELSEVIER, Expert Systems with Applications*, Vol. 28, pp. 127–135, 2005.

[6] K. J. Kim, “Financial time series forecasting using support vector machines”, *Neurocomputing,* Vol. 55(1/2), pp. 307–319, 2003.

[7] Emin Aleskerov, Bernd Freisleben and Bharat Rao, “CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection”, *Proceedings of the Computational Intelligence for Financial Engineering,* pp. 220-226, 1997.

[8] Zhang yongbin, You Fucheng and Liu Huaqum, “Behavior–Based Credit Card Fraud Detection Model”, *Fifth International Joint Conference on INC, IMS and* IDC, pp. 855-858, 2009.

[9] Chuang-Cheng Chiu and Chich-Yuan Tsai, “A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection”, *Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service*, pp. 177-181, 2004.

[10] Wen-Fang Yu and Na Wang, “Research on Credit Card Fraud Detection Model Based on Distance Sum”, *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 353-356, 2009.

[11] Tao Guo and Gui-Yang Li, “Neural Data Mining for Credit Card Fraud Detection”, *International conference on Machine Learning and Cybernetics*, Vol. 7, pp. 3630-3634, 2008.

[12] Suvasini Panigrahi, Amlan Kundu, Shamikr Sural and A.K. Majumadar, “Credit Card Fraud Detection: A Fusion Approach Using Dempster Shafer Theory And Bayesian Learning”, *Information Fusion,* Vol. 10, No. 4, pp. 354-363, 2009.

[13] Abhinav Srivastava, Amlan Kundu, Shamikr Sural and A.K. Majumadar, “ Credit Card Fraud Detection Using Hidden Markov Model”, *IEEE Transactions on Dependable and Secure computing,* Vol. 5, No. 1, pp. 37-48, 2008.

[14] Valdimir Zaslavsky and Anna Strizhak , “Credit Card Fraud Detection using Self Organizing Maps”, *Information & Security: An International Journal,* Vol. 18, pp. 48-63, 2006.

[15] Chen, R., Chiu, M., Huang, Y. and Chen, L. “ Detecting credit card fraud by using questionnaire-responded transaction model based on Support Vector Machines”, *Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning,* Vol. 3177, pp. 800-806, 2004.

[16] V. Dheepa and R. Dhanapal, “Analysis of credit card fraud detection systems”, *International Journal of Recent Trends in Engineering*, Vol. 2, No. 3, pp. 126-128, 2009.

[17] Dumais.S, “Using Support Vector Machines for text categorization”, *IEEE Intelligent Systems,* Vol. 13, No. 4, pp. 21–23, 1998.

[18] G. Dror, R. Sorek and S. Shamir, “Accurate identification of alternatively spliced exons using support vector machine”, *Bioinformatics*, Vol. 21, No. 7, pp. 897–901, 2005.

[19] Osuna. E, “Applying Support Vectors Machines to face detection”, *IEEE Intelligent Syst. Mag., Support Vector Machines,* Vol. 13, No. 4, pp. 23–26, 1998.

[20] Vapnik, V.N., “*The nature of statistical learning Theory*”, Springer, 1995.

[21] C. Cortes and V. Vapnik, “Support vector networks”, *Machine Learning*, Vol. 20, pp. 1-25, 1995.

[22] Cristianini N and Shawe-Taylor J, “*An Introduction to Support Vector Machines and other Kernel-based Learning Methods*”, Cambridge University Press, Cambridge, UK, 2000.

[23] Xuchen Li, Lei Wang and Eric Sung, “Ada Boost with SVM–based component classifiers”, *Engineering Applications of Artificial Intelligence*, Vol. 21, No. 5, pp. 785-795, 2008.

[24] Lindsay I Smith, “*A Tutorial on Principal Component Analysis*”, Feb 26, 2002.

[25] Chih-Wei Hsu, Chih-Chung Chang and Chi-Jen Lin, “A Practical Guide to Support Vector Classification”, Technical Report, National Taiwan University, Taiwan, 2010.

[26] Salvatore J Stoflo, David W Fan, Wenke Lee and Andreas L Prodromidis and Philip K Chan, “Cost –Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project”, *Proceedings of the DARPA Information Survivability Conference and Exposition,* Vol. 2, pp. 130-144, 2000.