An Optimized SVM Model for Detection of Fraudulent Online Credit Card Transactions

Xu Wei
School of Information and Safety Engineering
Zhongnan University of Economics and Law
Wuhan, China
aixilily@163.com

Liu Yuan
School of Information and Safety Engineering
Zhongnan University of Economics and Law
Wuhan, China
aixilily@163.com

***Abstract*—In order to identify the credit card fraudulent transactions, in this paper we propose an optimized SVM model for detection of fraudulent online credit card model. The model use non-liner SVM and RBF for the sparse transaction data, and use grid algorithm to determine the optional combination of parameters. There dose not exist local minima problem, the curse of dimensionality, small sample size problem and nonlinear problem. In the experiment of the simulation environment, the proposed model performance better than other models which verifies its feasibility.**

Keywords-SVM; fraud detection; credit card

I. INTRODUCTION

The development of economic and the open financial market make the credit card business become one of the bank's most important incomes. But along with the growth of issuance volume, global credit fraud transactions increase at an alarming rate. Financial companies can not effectively discover fraudulent transactions; as a consequence the loss is becoming increasingly serious. How to identify the credit card fraudulent transactions effectively, quickly and accurately is becoming the generally concerned problem.

In China, we begin to use a credit card of payment online in recent years. The related study is divided into two directions: fraudulent identification and enterprise applications. The researches of the first direction are Tong Fengru's which is based on the combination of classifier and Yan Hua, Hu Mengliang's which is using Bayesian classification algorithm. In the latter direction, the third party payment merchant—IPS officially released a credit card anti-fraud system called ANT [1]. The system uses a neural network-based anti-fraud model for parallel docking with the credit card payment instruments, effective inhibition of the credit card electronic payment to the various risks that may occur during the transactions.

Early research of credit card fraud prevention focuses on the classification and identification of methods and models, including single pattern recognition methods such as decision trees and neural networks, the combination method, distributed data mining. However due to the complexity and sparse of the transaction data, these methods are often faced with the issue of model selection, model parameter settings, improper selection when dealing with large scale transaction data, which often lead to owe study, over fitting and the local optimal problem[2]. Support vector machine is a relatively new field in the data mining field. The process is first mapped data from the input space to feature space, and then construct a linear discriminate function in feature space. Although there are many similarities between neural network and support vector machine in structure, the latter is relatively simple.

In this paper, we apply the support vector machine algorithm to construct an optimized SVM model for detection of fraudulent online credit card transaction, which helping the merchants make decision on weather to accept the deal. And then we analyze the test results of each model. The rest of the paper is organized as follows: The section 2 focuses on the characteristics and principles of support vector machine algorithm. Section 3 focuses on the choice of vector machine model. Section 4 gets the best support vector model after data processing, and then compares the test results with other studies to verify the feasibility of the model. Section 5 is the conclusion.

II. SUPPORT VECTOR MACHINE

A. Characteristics of the Support Vector Machine (SVM)

Support Vector Machine (SVM) is a popular machine learning method for classification, regression, and other learning tasks. The basic idea of SVM method is: the definition of the optimal linear hyper plane, and the search algorithm for optimal linear hyper plane by solving a convex programming problem. Then based on Mercer nuclear expansion theorem, through a nonlinear mapping φ, the sample space is mapped to a high-dimensional and even infinite dimensional feature space (Hilbert space), so that in the feature space can be applied to solve the linear learning machine method, the sample space, the highly nonlinear classification and regression problem.

This principle is based on the fact that: the error rate of SVM depends on the sum of training error and an item relies on VC dimension. In the case of classification, support vector machines make the previous value a value of zero, and make the second minimum. So SVM has good generalization performance in pattern classification.

B. Support Vector Machines

The core idea of SVM is the kernel function satisfies Mercer conditions instead of a non-linear mapping, so that the sample points in the input space can be mapped to a

978-0-7695-4853-1/12 $26.00 © 2012 IEEE
DOI 10.1109/ICMeCG.2012.39

IEEE computer society

high-dimensional feature space and linearly separable in the space, then construct an optimal hyper plane to approximate the ideal classification results, to avoid the curse of dimensionality in the feature space.

In the nonlinear case, the hyper plane is

$$w \cdot \phi(x) + b = 0$$

The decision-making function is

$$g(x) = \mathrm{sgn}(w \cdot \phi(x) + b)$$

Optimal hyper plane is described as

$$\min_{w,b,\xi i} \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

$$s.t., y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, 2, ..., N$$

The dual optimization problem is

$$\min_{\alpha} L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \phi(x_i) \phi(x_j)$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

$$s.t., 0 \leq \alpha_i \leq C, i = 1, 2, ..., N$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

where $K(x_i, x_j) = \phi(x_i)\phi(x_j)$ is the kernel function. Decision-making functions and parameters b are given respectively

$$g(x) = \mathrm{sgn}(\sum_{i=1}^{N} y_i \alpha_i K(x_i, x) + b) \quad (1)$$

$$b = \frac{1}{N_{NSV}} \sum_{x \in NSV} (y_i - \sum_{x_j \in SV} \alpha_j y_j K(x_j, x_i)) \quad (2)$$

According to (1) and (2), we can see that solving optimization problems and decision-making function dose not need to calculate the nonlinear function. We only need to calculate the kernel function, which avoids the curse of dimensionality of the feature space. The final discrimination function contains only seeking support vector inner product and its sum, thus the computational complexity will depend on the number of the SV [3].

The core issue in SVM research is the choice of the kernel function. However there is not more feasible and efficient method to construct a suitable kernel function in allusion to specific problem. In practice, the more commonly used kernel function has the following four categories [4].

- Linearity kernel function: $K(x, y) = x^T \cdot y$
- Multinomial kernel function:

$$K(x, y) = [x \cdot y + c]^d$$

- RBF (Radial Basis Function) kernel function:

$$K(x, y) = \exp\{-\|x - y\| / 2\sigma^2\}$$

- Sigmoid kernel function:

$$K(x, y) = \tanh(v(x \cdot y) + c)$$

III. Support Vector Machine Model Selection

If you want to apply SVM to specific problems, you need to select a kernel function. Although theoretically those functions satisfying the Mercer condition can be selected as the kernel function, classification performance is completely different when using different kernel function. Thus, for a specific problem, selecting a specific function is very important. And even a certain type of kernel function is selected; we need to choose the corresponding parameters, such as the polynomial order and the width of the radial function. Quadratic programming parameters such as C can also affect the classifier's generalization ability. Model selection includes the kernel function parameter selection, category selection and quadratic programming parameter selection. Although the research on model selection is not a lot, it is paid more attention by researchers as an important topic.

The grid search method is often used to determine the best model. Its basic idea is to remove the value of a number of model parameters respectively, and combine them into a number of combinations, and then train the SVM, estimate the accuracy of its learning, and finally find a combination of the most learning accuracy as the optimal combination of parameters. This topic uses the radial basis kernel function and the grid method to determine the optimal parameters (C, σ). The following is the specific process.

- Set the range and step size of (C, σ). (C, σ) generally take the value of the index type C range from 100 to 1500 step 100; value of σ is 0.5 to 8; the step size is 0.5. Then a two-dimensional grid is constituted in the coordinate system.
- The value of the grid in each group (C, σ) will be in accordance with the following method to calculate the accuracy of the prediction. Sample data is divided into training and test set. Test set is used test SVM classifier fostered by the training set. Support Vector Machine classification in the latter classification accuracy is treated as the actual performance of support vector machine on the unknown data.
- Finally, the accuracy of each group (C, σ) values is depicted by the contour lines, getting a contour map on the accuracy. According to SRM theory, the highest point in the figure is the most advantages and determines the optimal (C, σ) values.

The grid method has the advantage of parallel processing and SVM training is independent. Its disadvantage is the amount of calculation. Calculation in the case of two parameters is O (n2). Improved method based on structural properties of the SVM is: First, getting the value of optimal (C, σ) by using a larger step length. (C, σ is) combination. Then a more detailed grid search within a certain range next to the (C, σ) values. Then a more detailed grid search within a certain range is conducted next to the (C, σ) values.

IV. EXPERIMENT

A. Data preprocessing

The transaction data we used in this paper are from a commercial bank's business database. The database contains 2548 customers, including a total of 33562 transaction records in the last month of 2380 customers. The first step is the preprocessing of raw data and the main part is data cleaning and conversion.

1) data cleaning

a) Remove dirty data

Since there are 198 customers do not have transaction records, we need to remove the information of these customers from the raw data. And 128 records in the user table have transactions amounted to 0, they should removed as invalid records.

b) Handle missing values

The missing value is always the inevitable problems of each data set. Missing values can not be ignored, or mining results will be greatly affected. Usually there are several ways to deal with missing values: manual fill, mean fill, fill the default values, the most likely value of fill, fill with the class mean, and ignore the section record [5]. After removing the dirty data statistics, the number of null value in the data table is counted. The results are shown in TABLE 1.

TABLE I. MISSING VALUE TABLE

Field name	Missing items	Loss ratio
Family size	28	1.28%
Amount of the transaction	876	2.8%
Position	28	1.28%
Vocation	4	0.18%
Marital status	13	0.6%
Wage	5	0.23%

If "family size" field has missing values, we can count on family size field in the Customer Information Sheet. We can find that the proportion of 4 is the largest one, so we can use the value 4 filling in according to the law of default values. If the transaction amount field has missing values, mean imputation can be used. The average of the user's transaction volume can be used to fill in missing values. If other fields have missing values, we can use K-mean clustering method to fill the missing value with the center of each class.

c) Correct error data

To rectify the raw data which do not meet the business logic of the data values. Accordance with the principle of the credit card business to verify the data, some data not conform to business logic must be changed. Erroneous data in the raw data are mainly the following three.

- In age field, some values are less than 18 which are clearly inconsistent with the rules of age restrictions in the credit card business;
- "Credit line" field value is a floating-point or not credit standards, needing to use the value closest to the amount of value to correct;
- If "household population" field value is zero, we can refer to the "family size" field of missing values. After the completion of data cleaning, we will get a sample set of 20 attributes shown in TABLE 2 (Property is only part of the properties listed in the table).Please see TABLE 2 at the bottom of this column.

2) Data conversion

Data samples can not be applied directly after cleaning; you also need to convert data, conversion into numerical samples supported by vector machine. This dataset can be expressed by the following mathematical form.

$$((x_1, x_2, ..., x_{41}), y), 0 \le x \le 1, y \in \{-1, +1\}$$

"Transaction data" attribute need the attribute construction. After the attribute construction, it can be converted into the properties of the trading week.

B. Experimental data selection

Selecting training and testing data sets is a complex process in the establishment of a detection model for credit card online payment. Support vector machine is not sensitive for data distribution, and thus we should do our best to keep the original distribution of the data not change. It is also necessary to ensure that the training data set has adequate fraudulent transactions samples. We can observe from the composition of the sample there is a very serious category distribution

TABLE II. DATA CLEANING

Gender	Age	Family size	Marital status	Vocation	Company type	Line of credit	Education	Consumer date	The amount of consumption	Fraudulent trading
woman	48	3	unmarried	else	else	5000	Master degree	2011-07-07	74.10	1
woman	29	3	unmarried	teacher	state-run	3000	Primary school	2011-07-07	1850	1
woman	19	2	unmarried	student	else	3000	Undergraduate	2011-07-07	60	1
woman	56	3	unmarried	civil servant	state-run	5000	Master degree	2011-07-07	439	1
man	51	2	unmarried	service	individual	10000	Associate degree	2011-07-07	38.7	1
man	60	3	married	teacher	state-run	5000	Master degree	2011-07-07	49.1	1
woman	35	3	married	medical	individual	5000	Master degree	2011-07-07	150	1
woman	27	2	unmarried	art	individual	5000	Master degree	2011-07-07	1580	0

asymmetry between normal transactions and transactions of credit card fraud, and between normal customer and fraud customer. In the collected over 30,000 data samples, the proportion of normal trading and fraudulent transactions were 98.4 percent, 1.6 percent respectively. There are just over 500 fraudulent transactions samples.

C. Model Selection

After the training set is selected, SVM is used to find decision function. In this process, we need to select the SVM kernel function and its parameters. After comparisons, we choose the RFB function.

We use the previously described grid search method to test the classification performance of radial basis function. RBF function has two parameters: C and σ. We set the parameter C in the range 100 to 2000, the step length is 100; parameter σ ranges from 0.5 to 8, the step length is 0.5. A two-dimensional grid is constituted in (C, σ) coordinate system.

D. Experimental results and analysis

Credit card business data is often related to personal privacy. There is no public data set of a credit card transaction. The researches appear at home and abroad are obtained through a particular data set. The ID3+BP hybrid model is over fitting sometimes. Algorithm should be the main reason for this phenomenon. The over fitting will degrade the performance in the prediction set. The SVM model did not over fitting. However, there may be other reasons:

- Criteria. The ID3 + BP [6] criteria used by the hybrid model is different from this model. It uses two types of correct rate standard predictions, which may lead ultimately to the different selection of optimal model.
- Data mixed strategy. Using ID3 + BP hybrid model in the split of the experimental data, the mix of data strategy selected may lead to over fitting.

V. CONCLUSION

In this paper, on the basis of the risk faced by the credit card and online payment features, we focus on credit card online payment fraud risk detection and prevention. First starting from the support vector machine, we highlight the features and principles of support vector machine algorithm; then discuss the support vector machine model selection, and finally build the model of credit card fraud detection based on support vector machine. Then the model is applied to an anti-fraud system for credit card online payment. We do data preprocessing, find the best SVM model, and compare it with ID3+BP hybrid model. We find that the performance of the SVM model is higher than the performance of the ID3 + BP hybrid model and the former is not overly dependent on the user's "skills" during the experiment. As a result, we achieve the feasibility validation of the model.

ACKNOWLEDGMENT

This work is supported by National Social Science Foundation of China Project "Studies on Suspicious Financial Transaction Monitoring System based on Behavior Pattern Recognition" (09BTJ002).

REFERENCES

[1] Ming Xiao, The anti-fraud reserch for credit card online payment, Nankai University, May 2010.

[2] Pai, Ping-Feng, Chih-Shen Lin, A hybrid ARIMA and support vector machines model in stock price forecasting, Omega: International Journal of Management Science, 2005, 33(6): 497-505.

[3] Linhui Li, Intrusion Detection Based on Feature Selection, Zhongnan forestry university of science and technology, 2009.

[4] Ling Yang, Tongue color pattern recognition system, Nankai University, 2008.

[5] C. Chiu, C. Tsai: A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection[C].Proceedings of 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service,2004:177-181.

[6] Daqin Wei, Detection of risk of credit card transactions based on data mining, Chengdu: Sichuan Normal University, 2007.