Credit Risk Evaluation Using ES Based SVM-MK

Wei Liwei[1, a], Zhang Ying[1,b], Liu Mochen[2,c] and Xiao Qiang[3,d]

[1] China national institute of standardization, Beijing 100191, China

[2]Mechanical &electronic Engineering institute Shandong Agricultural University, Taian 271018,China

[3] Beijing International Electric Engineering Ltd.Co, Beijing 100039, China

[a]weilw@cnis.gov.cn, [b]zhangy@cnis.gov.cn, [c]liu_mochen@126.com, [d]aguyforeve@126.com

Keywords: Credit risk evaluation, SVM-MK, ES SVM-MK,SVM.

Abstract. Under the background of big data recent studies have revealed that emerging modern machine learning techniques are advantageous to statistical models for credit risk evaluation, such as SVM. In this study, we discuss the applications of the evolution strategies based support vector machine with mixture of kernel(ES based SVM-MK) to design a credit evaluation system, which can discriminate good creditors from bad ones. Differing from the standard SVM, the SVM-MK uses the 1-norm based object function and adopts the convex combinations of single feature basic kernels. Only a linear programming problem needs to be resolved and it greatly reduces the computational costs. A real life credit dataset from a US commercial bank is used to demonstrate the good performance of the ES SVM- MK.

Introduction

In the big data stage credit risk evaluation is an important field in the financial risk management. Extant evidence shows that in the past two decades bankruptcies and defaults have occurred at a higher rate than any other time. Thus, it's a crucial ability to accurately assess the existing risk and discriminate good applicants from bad ones for financial institutions, especially for any credit-granting institution, such as commercial banks and certain retailers. Due to this situation, many credit classification models have been developed to predict default accurately and some interesting results have been obtained.

These credit classification models apply a classification technique on similar data of previous customers to estimate the corresponding risk rate so that the customers can be classified as normal or default. Some researchers used the statistical models, such as Linear discriminate analysis [1], logistic analysis [2] and probit regression [3], in the credit risk evaluation. These models have been criticized for lack of classification precision because the covariance matrices of the good and bad credit classes are not likely to be equal.

Support vector machine (SVM) is first proposed by Vapnik[4]. Now it has been proved to be a powerful and promising data classification and function estimation tool. Reference [5] and [6] applied SVM to credit analysis. They have obtained some valuable results. But SVM is sensitive to outliers and noises in the training sample and has limited interpretability due to its kernel theory. Another problem is that SVM has a high computational complexity because of the solving of large scale quadratic programming in parameter iterative learning procedure.

Recently, the reference [7] draws the conclusion that the optimal kernel can always be obtained as a convex combinations of many finitely basic kernels. And some formulations [8],[9] have been proposed to perform the optimization in manner of convex combinations of basic kernels.

Motivated by above questions and ideas, we propose a new method named evolution strategies(ES) based support vector machines with mixture of kernel (ES based SVM-MK) to evaluate the credit risk. In this method the kernel is a convex combination of many finitely basic kernels. Each basic kernel has a kernel coefficient and is provided with a single feature. The 1-norm is utilized in SVM-MK. As a result, its objective function turns into a linear programming parameter iterative learning procedure and

greatly reduces the computational complexity. Furthermore, we can select the optimal feature subset automatically and get an interpretable model.

Support Vector Machine with Mixture of Kernel

Considering a training data set $G = \{(\vec{x}_i, y_i)\}_{i=1}^n$, $\vec{x}_i \in R^m$ is the ith input pattern and y_i is its corresponding observed result $y_i \in \{+1, -1\}$. In credit risk evaluation model, x_i denotes the attributes of applicants or creditors, y_i is the observed result of timely repayment.

The optimal separating hyper-plane is found by solving the following regularized optimization problem [6]:

$$\min J(\vec{w}, x) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n x_i \tag{1}$$

$$s.t. \begin{cases} y_i(\vec{w}^T f(\vec{x}_i) + b) \geq 1 - x_i \\ x_i \geq 0 \end{cases} \quad i = 1, \mathbf{L}\, n \tag{2}$$

where c is a constant denoting a trade-off between the margin and the sum of total errors. $f(\vec{x})$ is a nonlinear function that maps the input space into a higher dimensional feature space. The margin between the two parts is $2/\|w\|$.

The quadratic optimization problem can be solved by transforming Eq.1 and Eq.2 into the saddle point of the Lagrange dual function:

$$\max\left\{ \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j k(\vec{x}_i, \vec{x}_j) \right\} \tag{3}$$

$$s.t. \begin{cases} \sum_{i=1}^n y_i a_i = 0 \\ 0 \leq a_i \leq c, i = 1, \mathbf{L}, n \end{cases} \tag{4}$$

where $k(\vec{x}_i, \vec{x}_j) = f(\vec{x}_i) \bullet f(\vec{x}_j)$ is called the kernel function, a_i are the Lagrange multipliers.

In practice, a simple and efficient method is that the kernel function being illustrated as the convex of combinations of the basic kernel:

$$k(\vec{x}_i, \vec{x}_j) = \sum_{i=1}^n b_d k(x_{i,d}, x_{j,d}) \tag{5}$$

where $x_{i,d}$ denotes the dth component of the input vector $\vec{x}_i$.

Substituting Eq.5 into Eq.3, and multiplying Eq.3 and Eq.4 by b_d, suppose $g_{i,d} = a_i \cdot b_d$, then the Lagrange dual problem change into:

$$\max\left\{ \sum_{i,d} g_{i,d} - \frac{1}{2} \sum_{i,j,d=1}^{n,m} g_{i,d} g_{j,d} y_i y_j k(x_{i,d}, x_{j,d}) \right\} \tag{6}$$

$$s.t. \begin{cases} \sum_{i,d} y_i g_{i,d} = 0 \\ 0 \leq \sum_{d=1}^m g_{i,d} \leq c, i = 1, \mathbf{L}, n \\ g_{i,d} \geq 0, d = 1, \mathbf{L}, m \end{cases} \tag{7}$$

The new coefficient $g_{i,d}$ replaces the Lagrange coefficient a_i. The number of coefficient that needs to be optimized is increased from n to $n \times m$. It increases the computational cost especially when the number of the attributes in the dataset is large. The linear programming implementation of SVM is a promising approach to reduce the computational cost of SVM and attracts some scholars' attention. Based on above idea, a 1-norm based linear programming is proposed:

$$\min J(\vec{g}, \vec{x}) = \sum_{i,d} g_{i,d} + 1 \sum_{i=1}^n x_i \tag{8}$$

$$s.t. \begin{cases} y_i\left(\sum_{j,d} g_{j,d} y_j k\left(x_{i,d}, x_{j,d}\right) + b\right) \geq 1 - x_i \\ x_i \geq 0, i = 1, \mathbf{L}, n \\ g_{i,d} \geq 0, d = 1, \mathbf{L}, m \end{cases} \tag{9}$$

In Eq.8, the regularized parameter l controls the sparse of the coefficient $g_{i,d}$.

The dual of this linear programming is:

$$\max \sum_{i=1}^{n} u_i \tag{10}$$

$$s.t. \begin{cases} \sum_{i=1}^{n} u_i y_i y_j k\left(x_{i,d}, x_{j,d}\right) \leq 1, j = 1, \mathbf{L}, n. d = 1, \mathbf{L}, m \\ \sum_{i=1}^{n} u_i y_i = 0 \\ 0 \leq u_i \leq l, i = 1, \mathbf{L}, n \end{cases} \tag{11}$$

The choice of kernel function includes the linear kernel, polynomial kernel or RBF kernel. Thus, the SVM-MK classifier can be represented as:

$$f\left(\overset{\mathbf{v}}{x}\right) = sign\left(\sum_{j,d} g_{j,d} y_j k\left(x_{i,d}, x_{j,d}\right) + b\right) \tag{12}$$

It can be found that above linear programming formulation and its dual description is equivalent to that of the approach called 'mixture of kernel' [9]. So the new coefficient $g_{i,d}$ is called the mixture coefficient. Thus this approach is named "support vector machine with mixture of kernel" (SVM-MK). Comparing with the standard SVM that obtains the solution of a_i by solving a quadratic programming problem, the SVM-MK can obtain the value of mixture coefficient $g_{i,d}$ by solving a linear programming. So the SVM-MK model greatly reduces the computational complexity. It is more important that the sparse coefficients $g_{i,d}$ give us more choices to extract the satisfied features in the whole space spanned by all the attributes.

Experiment analysis

In this section, a real-world credit dataset is used to test the performance of SVM-MK. The dataset is from a major US commercial bank. It includes detailed information of 5000 applicants, and two classes are defined: good and bad creditor. Each record consists of 65 variables, such as payment history, transaction, opened account etc. This dataset includes 5000 applicants, in which the number of bad accounts is 815 and the others are good accounts. Thus the dataset is greatly imbalance. So we preprocess the data by means of sampling method and making use of 5-fold cross-validation to guarantee valid results. In addition, three evaluation criteria measure the efficiency of classification:

$$\text{Type } I \text{ error} = \frac{FP}{TP + FP} \tag{13}$$

$$\text{Type } II \text{ error} = \frac{FN}{TN + FN} \tag{14}$$

$$\text{Total error} = \frac{FP + FN}{n} \tag{15}$$

In the above three equations, TP = true observed good, FP = false observed good, TN = true observed bad, FN = false observed bad, n = the number of samples.

Evolution strategies (ES) for selection of the adaptive model

Based on the Darwinian principle of 'survival of the fittest', ES obtains the optimal solution after a series of iterative computations [10]. ES works with a set of candidate solutions called a population. The ES has three basic operations: mutation, recombination and selection.

As an optimization algorithm, ES generates successive populations of alternate solutions to the problem, until obtained acceptable results. A fitness function assesses the quality of a solution in the

evaluation process. Mutation and recombination functions are the main operators that randomly impact the fitness value. The evolutionary process operates for many generations, until the termination condition is satisfied.

It can be concluded that there are endogenous as well as exogenous strategy parameters in ES. Endogenous strategy parameters such as populations n of individuals a, specific object parameter set 1 and its fitness value F(1), can evolve during the evolution process, and are needed in self-adaptive ES. Strategy-specific parameters l and k, as well as q, are called exogenous strategy parameters which are kept constant during the evolution process.

To implement the proposed approach, this study uses the RBF kernel function for the SVM classifier because the RBF kernel function can analyze higher-dimensional data, which requires only two parameters, s^2 and l, to be defined.

When the RBF kernel is selected, hyper-parameters s^2 and l are used as input attributes that must be optimized using the proposed ES-based system. Therefore, the variable vector (chromosome) involves two variables for the main routines, like (x1, x2). In this study, we set q = 3, l = 3 and k = 7. We use the comma selection (l, k) to allow choice from the selection set which has forgotten the parents. This strategy relies on a birth surplus, i.e. on k > l, in a strict Darwinian sense of natural selection. And we use the overall hit rate as the fitness of ES.

Experiment result

Firstly, the data is normalized. In this method the Gaussian kernel is used, and the kernel parameter needs to be chosen. Thus the method has two parameters to be prepared set: the kernel parameter s^2 and the regularized parameter l. The Type I error (e1), Type II error (e2), Total error (e), number of selected features and the best pairs of (s^2, l) for each fold using ES based SVM-MK approach are shown in table 1. For this method, its average Type I error is 24.41%, average Type II error is 16.64%, average Total error is 22.76% and average number of selected features is 18.

Table 1.Experimental results for each fold using ES based SVM-MK

Fold#	e1 (%)	e2 (%)	e (%)	Optimized s^2	Optimized l	selected features number
1	20.2	15.7	18.9	4	5	19
2	27.95	15.19	26.1	5. 5	3	13
3	24.54	14.29	23.1	5	1	8
4	20.27	13.68	19.5	3	5	25
5	28.49	27.14	28.3	1.2	8	26
Average	24.25	17.24	23.22			18

The values of parameters to get the best prediction results are l = 3 and s^2=5.5, the number of selected features is 13, the bad ones' error is 15.19% and the total error is 26.1%, they are the lowest errors compared to the other situations. And almost eight of ten default creditors can be discriminated from the good ones using this model at the expense of denying a small number of the non-default creditors. A bigger parameter l results in selecting a good many of features. But the parameter s^2 has no effect on the feature selection. When the value of parameter s^2 matches the certain values of parameter l, we can get promising classification results. In general, there is a trade off between Type I and II error in which lower Type II error usually comes at the expense of higher Type I error.

Comparison of results of different credit risk evaluation models

The credit dataset that we used has imbalanced class distribution. Thus, there is non-uniform misclassifying cost at the same time. The cost of misclassifying a sample in the bad class is much higher

than that in the good class. There are only two different populations, the cost function in computing the expected misclassification is considered as follows:

$$\cos t = c_{21} \times p_1 \times p(2|1) + c_{12} \times p_2 \times p(1|2) \tag{16}$$

where c_{21} and c_{12} are the corresponding misclassification costs of Type I and Type II error, p_1 and p_2 are prior probabilities of good and bad credit applicants, $p(2|1)$ and $p(1|2)$ measure the probability of making Type I error and Type II error. In this study, $p(2|1)$ and $p(1|2)$ are respectively equal to Type I and Type II error. The misclassification ratio associated with Type I and Type II error are respectively 1 and 5 [11]. The priors of good and bad applicants are set to as 0.9 and 0.1 using the ratio of good and bad credit customers in the empirical dataset. In order to further evaluate the effectiveness of the proposed ES based SVM-MK credit evaluation model, the classification results are compared with some other methods using the same dataset, such as multiple criteria linear programming (MCLP), multiple criteria non-linear programming (MCNP), decision trees and neural network. The results of the four models quoted from the reference [12]. Table 2 summarizes the Type I, Type II and Total error of the five models and the corresponding expected misclassification costs (EMC).

Table 2.Errors and the expected misclassification costs of the five models

Model	e1 (%)	e2 (%)	e (%)	EMC
MCLP	24.49	59.39	30.18	0.51736
MCNP	49.03	17.18	43.84	0.52717
Decision Tree	47.91	17.3	42.92	0.51769
Neural Network	32.76	21.6	30.94	0.40284
ES based *SVM-MK*	24.41	16.64	22.76	0.30289

From table 2, we can conclude that the ES based SVM-MK model has better credit scoring capability in term of the overall error and the expected misclassification cost criterion in comparison with former four models. Consequently, the proposed ES based SVM-MK model can provide efficient alternatives in conducting credit evaluating tasks.

Conclusions

This paper presents a novel ES based SVM-MK credit risk evaluation model. By using the 1-norm and a convex combination of basic kernels, the object function which is a quadratic programming problem in the standard SVM becomes a linear programming parameter iterative learning problem so that greatly reducing the computational costs. In practice, it is not difficult to adjust kernel parameter and regularized parameter to obtain a satisfied classification result. Using the ES optimization, we have obtained good classification results and meanwhile demonstrated that ES based SVM-MK model is of good performance in credit scoring system. And we get only a few valuable attributes that can interpret a correlation between the credit and the customers' information. So the extractive features can help the loaner make correct decisions. Thus the ES based SVM-MK is a transparent model, and it provides efficient alternatives in conducting credit scoring tasks. Future studies will aim at generalizing the rules by the features that have been selected.

References

[1]G. Lee, T. K. Sung, N. Chang: Dynamics of modeling in data mining: Interpretive approach to bankruptcy prediction. Journal of Management Information Systems, Vol.16(1999), p.63-85.

[2]J. C. Wiginton: A note on the comparison of logit and discriminate models of consumer credit behavior. Journal of Financial Quantitative Analysis, Vol.15(1980), p.757-770.

[3]B.J. Grablowsky, W. K. Talley: Probit and discriminant functions for classifying credit applicants: A comparison. Journal of Economic Business. Vol.33(1981), 254-261.

[4]V. Vapnik: The nature of statistic learning theory. Springer, New York(1995).

[5]Y. Wang, S. Wang, K. Lai, A new fuzzy support vector machine to evaluate credit risk. IEEE Transactions on Fuzzy Systems, Vol.13(2005(6)), p.820-831.

[6]Wun-Hwa Chen, Jin-Ying Shih: A study of Taiwan's issuer credit rating systems using support vector machines. Expert Systems with Applications, Vol.30(2006), p.427-435.

[7]A. Ch. Micchelli, M. Pontil: Learning the kernel function via regularization. Journal of Machine Learning Research, Vol.6(2005), p.1099-1125.

[8]G. R.G. Lanckrient, N. Cristianini, P. Bartlett, L. El Ghaoui, M.I. Jordan: Learning the kernel matrix with semidefinite programming. Journal of Machine Learning Research, Vol.5(2004), p.27-72.

[9]F.R. Bach, G. R.G. Lanckrient, M.I. Jordan: Multiple kernel learning, conic duality and the SMO algorithm. Twenty First International Conference on Machine Learning (2004), p. 41-48.

[10]H.P. Schwefel, Evolutionsstrategie und numerische Optimierung, Dissertation, TU Berlin, Germany(1975).

[11]D. West: Neural network credit scoring models. Computers and Operations Research, Vol.27(2000), p. 1131-1152.

[12]L.W. Wei, W.W. Li, X. Qiang: A New Features Selection model: Least Squares Support Vector Machine with Mixture of Kernel. 5th International Conference on Advanced Engineering Materials and Technology (AEMT 2015), p.665-670.