Advances in Applied Economics and Finance (AAEF)
Vol. 1, No. 4, 2012, ISSN 2167-6348
Copyright © World Science Publisher, United States
www.worldsciencepublisher.org

Prediction Research of China's Import and Export Trade Based on Support Vector Regression

Ling Zhang, Chongjun Fan, Ziqiang Xu

Business School, University of Shanghai for Science and Technology, Shanghai 200093, China

Email: zlauct@sina.com

Abstract –The support vector machine (SVM) is a new type of statistical learning theory, forming a general a study machine based on the theory of structural risk minimization principle, featuring small sample and strong popularization ability ect, avoiding the 'dimension disaster', and has a good generalization ability. This paper uses SVM regression to predict China's import and export trade volume and in order to secure the accuracy, some index variables were brought in.The results show that SVM regression has very good prediction effect under the circumstance of small sample. Besides, some suggestions were given to improve the accuracy of trade volume prediction.

Keywords –Statistical learning; Support vector machine (SVM); Import and export forecast; Small sample

1. Introduction

Since the reform and opening up in 1979, China has made sustainable growth in import and export trade, by 2012 China has become the world's second largest trading nation and the first big export nation. In 2008, the financial crisis hit the whole world, China inevitably face the slump of the total import and export volume under the fact that the whole world's need shrink and the foreign trade once has two digits of descendent. Confronted the serious test, China's government take a serious measures to keep the steady development of foreign trade: improve the export tax rebates, reduce export premium rate, adjust the forbidden category of the processing trade, keep the basic stable of renminbi exchange rate. Since the fourth quarter in 2009, foreign trade gradually turned around the declining tendency. In 2010, the total amount of import and export trade volume is $2.97276 trillion, in 2011 the figure is $3.6407 trillion, increased by 18.3% compared with 2010. We can tell that the trade volume has shown the strong volatility and uncertainty under the influence of economic environment home and aboard, trade policy, emergency and so on. These factors have brought the difficulty in predicting of import and export trade volume accurately. The arrival of post financial crisis, the reshuffle of the world's economy requires a fresh survey of our trade economic environment. A better prediction of foreign trade will pose an important influence on the decision making process of foreign trade companies and even the ministry of commerce in this new period.

The foreign trade prediction problem is a time serious problem, import and export trade data is a complex time serious featuring strong volatility and obvious randomness. At present, domestic research is mainly about the traditional simple algorithm, such as traditional time sequence method, regression analysis etc. All these statistic algorithms are based on a large number of sample data, whereas the small sample is very hard to find out the statistical rules in it, so it's imperative to find an algorithm based on small sample。The Support Vector Machine (SVM) method is a kind of machine learning method based on structural risk minimization and VC dimension theory in statistical learning theory.SVM is also called support vector method featuring complete theory, strong adaptability, global optimization, short training time, good generalization ability, has become a hot research spot home and aboard. The core content of SVM hasn't been brought until 1992, SVM is by far the most successful implementation of statistical learning theory. SVM has a lot advantages in small sample learning, such as the generalization ability of the model is improved and will not worry about the 'over fitting' problem. For now, SVM model has been applied in the forecast of real estate [4], the prediction of product sales[5], prediction of financial time series etc.

This paper do a research on prediction of China's import and export trade quarterly data using the method of SVM, and has got relatively accurate prediction results, providing a new idea of the prediction of China's import and export trade volume.

2. The principle of support vector regression model

2.1 Support vector regression (SVR)

Support vector machine (SVM) method is brought out from the optimal hyperplane when the data is linearly separable. The optimal hyperplane will not only classify all the training sample correctly and will make the distance between the optimal hyperplane and the point which is the nearest point to the optimal hyperplane. To control the complexity of classified machine through the maximization of distance so as to achieve good generalization ability. If the data is linearly inseparable, there is the generalized optimal hyperplane,which is in pursuit of the maximization of classification intervals while minimize the number of the wrong classified points.

SVM is used to solve the problem of classification and pattern recognition, some researchers have introduce insensitive loss function and brought out SVM regression method. This paper is using $\varepsilon -$ support vector regression, ε is the insensitive loss function （ $\varepsilon \geq 0$ ） ， which can be used to form the Support Vector Regression (SVR) .

SVR is a convex quadratic programming problem under constraint, based on the Mercer core

expansion theorem, through the nonlinear mapping, mapping the sample data to high dimension space G, introducing the insensitive loss function to high dimension space G, defining the optimal linear regression hyperplane, changing the algorithm of finding the optimal linear regression hyperplane to the solvement of a convex quadratic programming problem under constraint, so the only solution is a global optimal solution.

Set the sample data (xi, yi), I = 1, 2,... n, xi ∈Rn, yi ∈ R, yi is the expected value, n is the sample capacity, with y = g (x) = (w. (x)) + b as the estimate function. The original machine optimization problem of the support vector regression (SVR) is actually:

$$MinS = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i * + \xi_i) \qquad (1)$$

Constraint:

$$\begin{aligned} & y_i - [(\omega \cdot x) + b] \leq \varepsilon + \xi^*{}_i \\ & [(\omega \cdot x) + b] - y_i \leq \varepsilon + \xi_i \\ & \xi^*{}_i, \xi_i \geq 0, i = 1,.....,n \end{aligned} \qquad (2)$$

C is the regularization parameters, controlling the punishment degree of going beyond the error samples. ξi, ξ^*i are the upper and lower limits of slack variable. In formula (1), the first term makes the distance between sample and hyperplane as much as possible, the second term makes the classification error as less as possible, at the same time, try to classify correctly make the classification gap as big as possible so we can have generalized optimal hyperplane.

After choosing the proper kernel function, suitable precision and regularization parameter C, we can solve the problem of convex quadratic programming problem under constraint.

$$\begin{aligned} MinS = & \sum_{i,j=1}^{n}(a_i * - a_i)(a_j^* - a_j)K(x_i, y_i) + \\ & \varepsilon\sum_{i=1}^{n}(a_i * + a_i) - \sum_{i=1}^{n} y_i(a_i * - a_i) \end{aligned} \qquad (3)$$

constraint：

$$\begin{aligned} & \sum_{i=1}^{n}(a_i * - a_i) = 0 \\ & 0 \leq a_i *, a_i \leq C, i = 1,...,n \end{aligned} \qquad (4)$$

So the support vector regression function is:

(5)

$$y = g(x) = \sum_{i=1}^{n}(a_i * - a_i)K(x_i, y_i) + b \qquad (5)$$

2.2 The choice of the kernel function

In the training process of the support vector, especially when the problem is linearly non-separable, the calculation of the inner product of the sample is a time-consuming work. The use of nuclear function can transfer the linearly non-separable sample data in the high dimension space into the linearly separable, which will avoid the dimension disaster problem in the process of calculation. This is another advantage that SVM is better than the traditional statistical learning theory.

The different algorithms in SVM use the different inner product kernel function, Only the inner product kernel function satisfy the Mercer condition, can it correspond the inner product in a feature space. Currently, the more used nuclear functions are the following types:

1 ）polynomial kernel function

$$K(x, x^{'}) = [(x, x^{'}) + 1]^d$$

2 ）radial basis function

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

3 ）Sigmoid kernel function

$$K(x, x^{'}) = \tanh[\upsilon(x \cdot x^{'}) + c]$$

Different questions using different kernel functions to carry on the forecast, after many researchers study, we can tell that the forecast ability of RBF kernel function is better than the other two. This paper is also using RBF kernel function to build the evaluation function of SVR to carry on the forecast.

3. The forecast of China’s import and export trade volume based on the SVR

3.1 Data source and data pretreatment

This paper used the data of import and export volume of our country from Jan 2001 to Dec 2011 to carry on the forecast, the data was conducted quarterly and the results were showed quarterly, the original data was showed in table 1.

Table 1. The quarterly import and export volume in 2001-2011 (unit: 100millnion$)

YEAR	1st quarter	2nd quarter	3rd quarter	4th quarter
2001	1133.7	1276.7	1353	1334.3
2002	1219.8	1486.4	1744.8	1756.7
2003	1736.3	2023.1	2301.4	2451.3
2004	2398.3	2838.9	3046.8	3263.7
2005	2952.4	3497.4	3793.7	3977.7
2006	3712	4246.2	4765.5	4883.3

2007	4576.7	5237	5892.8	6032.2
2008	5708.8	6639.7	7396	5946.4
2009	5705.9	5174.3	6108.5	6487.7
2010	6176.7	7178.6	7939.8	8237.9
2011	8000.5	9029.1	9724.7	9652.7

This paper uses the data from Jan 2001 to Aug 2010 as the training sample and the fourth quarter in 2010 to the fourth quarter in 2011 as the test sample. From table 1, we can see that China’s import and export data have obvious seasonal characteristics. The traditional time series prediction method is based on the historical data to find the changing pattern, but this method requires the data have a stable changing trend and the precision of the strong votality time series is poorer, such as the import and export volume in 2008 and 2009. To overcome the uncertainty and seasonal problems of our country’s import and export volume time series, this paper brought several relevant index variables to build the model. The relevant index variables are shown in table 2.

Table2. The input index variables

	Index variables
X1	Last quarter import and export volume
X2	Last two quarter import and export volume
X3	Last three quarter import and export volume
X4	Last four quarter import and export volume
X5	Average exchange rate between RMB and dollar
X6	Average CPI
X7	foreign trade dependency
X8	Quarter index

Among them, X5，X6，X7，X8 are processed from the quarter data, X8-quarter index indicates the number of the four season, because we can see the obvious seasonal fluctuation of the import and export volume, normally, the first quarter’s volume is poorer, the next three quarters will grow step by step, in the fourth quarter the volume is the biggest in one year.

Because of the different index dimensions and value range, and different index variables will somehow influence the accuracy of the prediction, so the standardized processing of the data is a very necessary step before prediction. This paper uses the commonly used method-maximize and minimize standardized to scale the original data, changing the data into the value between 0 and 1.

x’=(x-MinValue)/ (MaxValue-MinValue)

x,x’ are the values respectively before and after the conversion, MaxValue is the biggest value of the sample, MinValue is the smallest value. The method- maximize and minimize standardized can maintain the relations of the original data.

3.2 Results and analysis of the forecast

The training software this paper is using if LIBSVM, which is developed by Dr. Chih-ChungChang from Taiwan University, it’s a tool software of SVM which can be used in the classification and regression of support vector machine and it also provides four commonly used kernel functions. LIBSVM is an open source package, it gives both the source code and the executable file under windows operating system, and at the same time users and also improve the algorithm according to the needs.

This paper uses the python language to find the parameters crossly, the parameters is C=1024, g=0.0625,p=8,MSE=0.02, MSE is used to evaluate the quality of the regression model. Finding the best parameters is a time consuming work, a small wave of the parameters can bring great flounce which leads to the multiple trials.

After selecting the parameters, we can build the model of the training samples and predict the test samples. Table 2 shows the results of SVR prediction and the comparison between SVR and the traditional time series analysis model-ARMA.

Table3. Comparisons of the two prediction models (unit: 100millnion$)

Year	volume	SVM predict results	relative error	ARMA predict results	relative error
2010 4th quarter	8237.9	8268.8	0.30%	7790.4	5.40%
2011 1st quarter	8000.5	7839.9	2.00%	7623.3	4.70%
2011 2nd quarter	9029.1	8699.7	3.60%	8198	9.20%
2011 3rd quarter	9724.7	968.2	0.38%	9889.9	2.70%
2011 4th quarter	9652.7	9810.1	1.60%	11022.3	14.10%

The evaluation of the prediction accuracy of SVR and ARMA is shown in table 4.

Table 4. Average relation error

	Average relative error
SVM prediction	1.5%
ARMA prediction	7.2%

From table 2 and table 3, SVM can predict our country’s import and export volume with relatively high accuracy, while the ARMA predict ability is not as good as SVM. We can tell that SVM has a great advantage in predicting the small sample of nonlinear time series.

4. Conclusions

We can see from the prediction results, SVM has a good prediction ability of complex import and export trade data of small sample and strong volatility. In this paper, the parameters are found through the popular way – cross inspection. The return of SVM forecast, the selection of parameters is through the current relatively .The selection of parameters directly affect the accuracy of prediction results. Sometimes the parameters that the cross inspection found are the optimal ones, we need to adjust the parameters according to the actual needs, which will need the patience and experiences. Therefore, how to find a method to find the optimal parameters still need further research.

As far as the accuracy of the prediction, there are ways to improve it. On one hand, in view of the particularity of import and export time series , we can select more index variables as the training data in order to find more factors that influence the trade volume. On the other hand, the way that preprocess the original data somehow influence the accuracy of the prediction, find a proper way to preprocess the data according to the actual need is very necessary.

SVR prediction is not a very mature prediction model, it is influenced by the selection of the parameters, original data, and the speed of the calculation is becoming slower when the sample is becoming bigger, all these will need further study.

References:

[1] Vladimir n. Vapnik the nature of statistical learning theory [M]. ZhangXueGong, al. Beijing: tsinghua university press, 2000.

[2] TRAFAL IS T B, INCO H. Support vector machine for regression and applications to financial forecasting [C].Proceedings of the IEEE - INNS - ENNS International Joint Conference on Neural Networks .Como, Italy, 2000, 6: 348 – 3531.

[3] Liong S Y, Sivapragasm C.F lood stage forecasting with SVM [J]. Journal of the American Water Resources Association, 2002, 38 (1) : 173-186.

[4] Li Chao, Fan Chongjun. Predition of model of chaotic time series based on support vector machines and its application to the real estate market [J]. Mathematics in Practice and Theory, 2011, (19).

[5] Du xiaofang, Zhang Jinlong. A Support Vector Machine Method for Sales Forecast of Farm Products [J]. Chinese management science, 2005, 13(4).

[6] Zou Baixian, Liu Qiang. Network flow prediction based on the ARMA model [J]. Journal of computer research and development, 2002, 39 (12).

[7] Zhao Jie. China's trade surplus forecast and outlook analysis based on the ARMA model [J]. China business, 2011 (25).

[8] Huang Yuansheng, Zheng Yan, Qi Jianxun. Power demand forecast based on least square support vector machine [J]. China management science, 2005, 13 (zl).

[9] John C . Platt . Sequential Minimal Optimization : A Fast Algorithm for Training Support Vector Machines [R] . Technical Report MSR—TR□ 98—14，April 21，1998．

[10] Zhou Wanlong, Yao Yan. Short-term forecast of stock price based on the support vector machine [J]. Business research, 2006, (6).

[11] Li ZhiLong, Chen Zhigang, Qin Zhi. Tourism demand forecast based on support vector machine [J]. Economic geography, 2010, 30 (12).

[12] Chen Ke, Ke Wende. Oil price forecast based on the support vector machine [J]. Computer simulation, 2011, 28 (12).

[13] Ji Aibing, Sun Jianping, Pang Jiahong. Combined forecast methods and the application based on support vector machine (SVM) f [J]. Statistics and decision-making, 2005 (5).

Vitae

Zhang Ling, was born in 1987. She is now a postgraduate Student in University of Shanghai for Science and Technology.

Fan Chongjun, was born in 1963. He obtained his postdoctoral degree in Shanghai Jiao Tong University. He is now a professor in Business School of University of Shanghai for Science and Technology.

Xu Ziqiang, was born in 1987. He is now a postgraduate Student in University of Shanghai for Science and Technology.