http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (16): 3660-3664, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3660.3664

Rmb/usd Exchange Rate Prediction Based on Support Vector Machine and Principal Component Analysis

Ling Wei and Wei-Xi Zhang
School of Management, Harbin University of Science and Technology, Harbin, 150000, Peoples Republic of China, China

Abstract: RMB/USD exchange rate bridges the economic exchanges between the China and America. And the prediction of the exchange rate has important practical significance for the enterprises, organizations and individuals. With the exchange rate theory and support vector machine theory as the main theoretical basis, a new index system and a model for the prediction of the exchange rate is constructed based on the principal component analysis and support vector machines. In order to verify the model effects, the results of the empirical study are compared with those of Genetic Algorithm-Backpropagation model and the traditional support vector machines. The more significant results of the proposed model illustrate that the selected indicators can accurately reflect the change factors of the exchange rate, that the principal component analysis can not only reduces the inputs for SVM but also improve the prediction accuracy and that the constructed model has obvious stable advantages, esp. with a small amount of samples.

Key words: Exchange rate prediction, support vector machine, principal component analysis, exchange rate theory

INTRODUCTION

In the process of economic globalization, the exchange rate, as the relative price of two currencies, is an increasingly important bridge and link in maintaining the economic relations between two or more countries. The exchange rate is a core economic variable under the open economy condition; it links a variety of macroeconomic factors and micro-economic factors and affect the national economy's internal balance and external balance. So, exchange rate forecasting has important practical significance for countries, businesses, researchers and other economies organizations. Therefore, with the exchange rate determination theory, a model for the exchange rate prediction is constructed based on Support Vector Machine (SVM) and Principal Component Analysis (PCA).

LITERATURE REVIEW

The exchange rate prediction researches are conducted mainly in two ways: fundamental analysis and technical analysis. The fundamental analysis emphasizes that the exchange rate changes are driven by internal factors, such as political, economic and unexpected factors while the technical analysis, usually used for the exchange rate prediction in a short period (within a day or each day), predicts the exchange rate directly using exchange data, Stochastic Oscillator (KD), Moving Average (MA), Relative Strength Index (RSI), etc. In the exchange rate prediction, the mostly used models are traditional statistical models, nonparametric models and the integrated model made up of different methods. Grossmann and Simpson (2010) used the linear prediction technique based on the relative Purchasing Power Parity (PPP) and achieved better results compared to those of the random walk model. Baghestani (2009) used the random walk model to predict the exchange rate between the U.S. dollar and the euro dollar from 1999-2007 and the exchange rate between the U.S. dollar and pounds from 1971-2007. The research results indicated that the random walk model could make an unbiased and accurate prediction of the change direction of the exchange rate. Carriero *et al.* (2009) pointed out that in the exchange rate prediction, the prediction model based on the economic theory was not better than the univariate random walk model in terms of the prediction results, especially in the short-term prediction.

SVMs, a group of supervised learning methods, can be applied to solve various real world problems in classification and regression. Currently, the mostly used method to predict the exchange rate is the integrated

Corresponding Author: Ling Wei, School of Management, Harbin University of Science and Technology, Harbin, 150001, Peoples Republic of China, China

model. Among the various methods to predict the exchange rate, non-linear neural network or SVM obtain higher prediction accuracy and these two methods are currently more used with economic indicators. On the theoretical and applied research of SVM, there is considerable room for improvement mainly in two aspects. On one hand, wavelet analysis (Lan *et al.*, 2012) and PCA (Subasi and Gursoy, 2010) are used to preprocess the data for SVM model, optimize the inputs and improve the final prediction results; on the other hand, SVM itself is improved, including the Integration with fuzzy theory (Wu *et al.*, 2009) the improved SVM calculation (Cheng *et al.*, 2013) parameter optimization (Ranaee *et al.*, 2010) the improved kernel function (Nelson *et al.*, 2008) and so on.

METHODOLOGY

Construction of index system: According to the principles of rationality, objectivity, comprehensiveness and data availability, the factors of the index system to build a sound and reliable model contain GDP, interest rates, CPI, Broad Money (M2), foreign exchange reserves, gold reserves, the consumer confidence index, total imports, total exports and the reference exchange rate. The other nine indicators except the reference rate all include the data from China and U.S.

Process of PCA-SV m model:

- **Data preprocessing:** The data preprocessing includes two aspects: the estimation of monthly GDP and data normalization

U.S. releases quarterly GDP data while China release GDP data each quarter but in a cumulatively way. So, China's quarterly GDP first needs to be calculated as follows.

$$GDP_q = GDP_{m1}+GDP_{m2}+GDP_{m3} \quad (1)$$

$$GDP_{m2} = GDP_{m1}\times(1+a) \quad (2)$$

$$GDP_{m3} = GDP_{m2}\times(1+a) \quad (3)$$

where, C is the growth rate of GDP in the current quarter, GDP_a is the GDP in the current quarter, $GDP_{m1}$is the GDP of the first month in the current quarter, GDP_{m2} is the GDP of the second month in the current quarter and GDP_{m3} is the GDP of the third month in the current quarter.
With the known GDP_a and C, GDP of each month in a certain quarter can be calculated using Eq. 1-3.

The reference exchange rate will be normalized separately before SVM process. The method of mean standard deviation is used to normalize data and the reference exchange rate with Eq. 4:

$$x\frac{1}{2}=\frac{x_{2-x_{m\theta an}}}{x_{m\theta an}} \quad (4)$$

where, $X_{m\theta an}$ is the mean of the data sequence and X_{var} is standard deviation of the data sequence.

- **Integration of PCA and SVM:** Let $X = (X_1, X_2, ...,X_{18})$ represent the 18 indicators to predict the exchange rate, fac(x) the PCA, $F = (f_1, f_2,..., f_n,n<18)$ the principal components obtained through PCA, S(x) the process of SVM, y the final prediction result and the model can be expressed in Eq. 5:

$$y = S(F) = S\ (fac(x)) \quad (5)$$

- **Parameter selection for SVM:** This study applies the most frequently used algorithm of is Support Vector Regression (SVR), ε-SVR and the best performed Radial Basis Function (RBF) as the kernel function. C, δ and ε are determined with the Grid Traversal Algorithm (GTA). The so-called GTA is to first determine the variation ranges and step sizes of the changes for, δ and ε. And then test each group of values in this range. Eventually, choose the best set as the parameters for the final training. The parameters' effects can be effectively improved through gradually narrowing the range and steps and the repeated GTA. The approach of k-fold cross-validation is used to determine the optimal parameters

Evaluation indicators of the model prediction accuracy.

The prediction accuracy of the model can generally be reflected by the prediction error with Mean Squared Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) whose calculations are as follows. The smaller the prediction error is, the more accurate the prediction model is:

$$MSE = \frac{\sum e_i^2}{n} \quad (6)$$

$$MAE = \frac{\sum |e_i|}{n} \quad (7)$$

Appendix 1: Variances of the principal components with PCA

Principal component	Initial eigenvalues Total	Percentage of variance	Cumulative (%)	Extraction of square and loads Total	Percentage of variance	Cumulative (%)
1	10.704	59.469	59.469	10.704	59.469	59.469
2	4.049	22.492	81.961	4.049	22.492	81.961
3	0.903	5.018	86.979	0.903	5.018	86.979
4	0.817	4.537	91.516	0.817	4.537	91.537
5	0.576	3.200	94.715	0.576	3.200	94.715
6	0.240	1.332	96.048	0.240	1.332	96.084
7	0.194	1.078	97.126	0.194	1.078	97.126
8	0.170	0.946	98.072	0.170	0.946	98.072
9	0.108	0.598	98.670			
10	0.079	0.438	99.109			
11	0.062	0.343	99.452			
12	0.037	0.206	99.658			
13	0.025	0.140	99.797			
14	0.017	0.092	99.889			
15	0.009	0.050	99.939			
16	0.006	0.034	99.973			
17	0.004	0.023	99.996			
18	0.001	0.004	100.000			

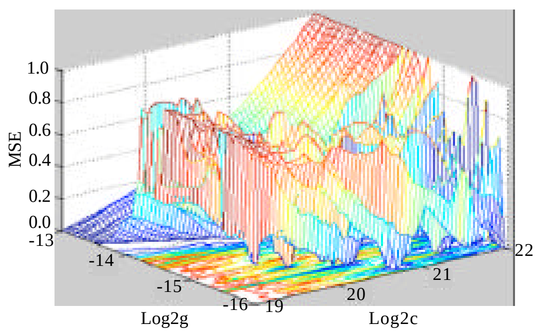

Fig. 1 3D map of SVR parameter optimization

$$\text{MAPE} = \frac{\sum |e_i / y_i|}{n} \tag{8}$$

where e_i is the value of error; n is the number of the predicted value; y_i is the true value corresponding to the predicted value.

EMPIRICAL STUDY

Sample selection and data sources: The exchange rate between and Renminbi (RMB, Chinese currency) and U.S. Dollars (USD) has a more substantial change after July, 2005 when People's Bank of China issued a proclamation declaring that China began to implement the floating exchange rate system. Therefore, 67 sets of monthly data from July 2005 to January 2011 are selected as the samples of the exchange rates. Among them, 52 sets of data from July 2005 to October 2009 are used as the training samples and 15 sets of data from November 2009 to January 2011 as the predictive samples. The data come from http: //www.safe.gov.cn/model_safe/tjsj/tjsj_detail.jsp?ID = 111100000000000000, 6 and id = 5; http://data.caing.com / macro / # top and the International Financial Statistics database (IFS).

Empirical study of PCA-SVM: The empirical research is conducted with SPSS16.0, Matlab R2010a and libsvm-mat-2.89-3 [FarutoUltimate3.0 Mcode]:

- **Extraction of the principal components:** After PCA preprocessing, the first eight principal components represent 98.07% variable information. Therefore, this study uses the eight principal components to represent the 18 initial variables. In the default setting, SPSS uses the correlation matrix for PCA with the following equations:

$$f_1 = 0.0081\text{gdbc}-0.033\text{rc}+0.009\ \text{cpic}+\cdots+0.016\text{imu}+0.071\text{exu} \tag{9}$$

$$f_2 = 0.008\ \text{gdbc}+0.192\text{rc}+0.232\ \text{cpic}+\ldots+0.238\text{imu}+0.155\text{exu} \tag{10}$$

$$\ldots$$

$$f_8 = 0.688\text{gdbc}-0.546\text{rc}+1.147\ \text{cpic}+\ldots+0.283\text{imu}-0.263\text{exu} \tag{11}$$

Since the variables gdpc, rc, cpic in Eq. 9-11 are normalized values, the value of 8 principal components can be calculated using these equations or SPSS. See Appendix 1.

- **Parameter optimization with SVM:** To process the data with SVM, GTA is used to determine the penalty factor C, the kernel parameter δ and the parameter of loss function ε with the optimization function SVMcg

For Regress of LIBSVM. For the easy explanation and the consistence with LIBSVM, g replaces δ, pr eplaces ε. Since there are only 52 training samples, to select the optimal parameters, let k = >in the K-fold cross- validation. To determine the range of C and g, start with a relatively large range, i.e. the range of C is[2][^{15,225}], g [2][$^{-18,2-8}$] . The step size of C and g is the default step size of SVMcg For Regress 0.8 i.e., the multiplication of each step size of C and g and 2^{0-8}. Thus, the relatively optimal C and g are obtian ed. Then, narrow the range of C and g. Let the range of C, g and the step size 0.0-5 .(i.e. each change of C is the multiplication of the initial value and g.) The value of 2^{0-05} and C is determined as g The value of p is determined with trial and error method as.

C = 2246771,86050491 and
g = 0000388965035192645

- **Training and prediction results of SVM:** After the determination of the parameter C, g and p train the prediction model PS-model in Matlab, using the training command svmtrain of LIBSVM. Predict the exchange rate from November 2009 to January 2011 using the prediction model and prediction command svmpredict. Using Eq. 6-8 and error values, the MSE, MAE, MAPE values of the PCA-SVM model can be calculated as7.7024, 2.5662 and 0.0038. As can be seen, the minimum error PCA-SVM model to predict the exchange rate for each period is 1.01 and the maximum is 4.63 with a mean of 2.5662. There are no values with obviously larger error than the other group which shows the prediction results of the exchange rate are relatively stable. In addition, the percentage of the maximum error is 0.0068 and the percentages of all the errors are less than 0.7% with MAPE of 0.0038, fully illustrating the high prediction accuracy of PCA-SVM model

Comparison between SVM and GA-BP model:

- **Prediction of the exchange rate based on SVM model:** Because SVM model does not process the raw data through PCA, each index has great difference in dimensions. Therefore, it is necessary to normalize the input indicators with the maximum and minimum method, mapping the input factors into [-1, 1]. Then optimize C and g with grid optimization with RBKF. In the K-fold cross-validation, k = 7. Parameter p is still determined using the trial and error method to obtain C = 2088204013179987, g = 0.009618315729257, p = 0.000009. Finally, train and predict with the model. The value of MSE, MAE, MAPE respectively is 31.0586, 4.9613 and 0.0073. With a relatively large error span, the minimum error is 1.45 and the maximum is10.52 which means that the prediction results of the exchange rate based on SVM model, to some extent, are not stable enough. Out of the total fifteen sets of data, the error percentage of four sets is more than of 1% and the highest gets to 0.0154. The error is relatively large compared with PCA-SVM model
- **Prediction of the exchange rate based on GA-BP model:** Before the application of GA-BP model, the data must be normalized. By contrast verification, the model is found to have the highest prediction accuracy when the input factors and the output exchange rate are normalized to [-1, 1] with the maximum and minimum method. Therefore, the input and output data of GA-BP model are mapped within this interval

In this study, the trial and error method is used to determine the parameters for GA-BP model. After several contrasts, the best set of parameters is eventually adopted. In addition, after the comparison of default algorithm Lenvenberg-Marquardt of BPNN in Matlab and Bayesian Regulation with the function trainbr in Matlab, Bayesian Regulation is used for its relatively better prediction effects.

Because of the randomness in the GA-BP process, in the variation of GA process and in the training process BPNN, the final outputs of training samples based on the GA-BP model are substantially inconsistent even though the same parameters are used each time. To overcome the randomness, 30 training and prediction are conducted and GA-BP model And choose the set with the best prediction effects.

The MSE, MAE, MAPE value of GA-BP model is 21.1745, 3.8745 and 0.0057, respectively. Among the fifteen sets of data, the minimum error is 0.15 and the maximum error is 7.61. The range of error values is slightly larger than that of the PCA-SVM model but smaller than that of SVM model; The error percentage of three sets exceeds 1% with the maximum value of 0.0111 which is also between the PCA-SVM model and the SVM model; In addition, in the 45 sets of data of the 15 months, the error of the GA-BP model is the smallest with the error value of 0.15 and the error percentage of 0.02%.

- **Comparison of the prediction results:** The predicted value of the exchange rate obtained using the PCA-SVM, SVM and GA-BP models are shown in Table 1

Table 1: Comparison of the prediction results of the PCA-SVM, SVM and GA-BP models

Month	Actual value	Predicted value of PCA-SVM	Predicted value of SVM	Predicted value of GA-BP
Nov. 2009	682.74	681.30	679.22	679.52
Dec. 2009	682.79	680.39	676.24	679.56
Jan. 2010	682.73	680.63	677.11	678.15
Feb. 2010	682.70	680.06	677.19	675.10
Mar. 2010	682.64	678.99	672.12	675.83
Apr. 2010	682.62	679.76	675.22	679.05
May 2010	682.74	679.52	680.66	678.41
Jun. 2010	681.65	679.81	680.31	675.34
Jul. 2010	677.75	678.98	680.53	678.47
Aug. 2010	679.01	674.39	680.46	671.92
Sep. 2010	674.62	675.63	678.24	674.77
Oct. 2010	667.32	671.38	673.19	673.45
Nov. 2010	665.58	662.67	669.12	665.34
Dec. 2010	665.15	661.91	672.01	665.96
Jan. 2011	660.27	661.52	668.02	663.59

Table 2: Comparison of the PCA-SVM model, the SVM model and the GA-BP model

Model	MSE	MAE	MAPE
PCA-SVM	7.7024	2.5662	0.0038
SVM	31.0586	4.9613	0.0073
GA-BP	21.1745	3.8745	0.0057

From the analysis of the predicted results of three models, we can get the preliminary conclusions, i.e., the PCA-SVM model predictions are the closest to the actual values of the exchange rate value. Moreover, the PCA-SVM model has a better prediction effect on the exchange rate value as well as the overall prediction trends and change amplitude; GA-BP model has a somewhat worse effect than the PCA-SVM model but slightly better the SVM model. For a more accurate comparison of the three models (Table 2).

As can be seen, first, the selected indicators can accurately reflect the change factors of the exchange rate. Sec, the prediction effect of the PCA-SVM model is significantly better than that of the SVM model. These results suggest that when the number of principal components is correctly retained, PCA can not only reduces the number of inputs for SVM but also improve the prediction accuracy. Third, the empirical analysis of the GA-BP model shows a low stability. The comparison results show that with a small number of samples, the PCA-SVM model has obvious advantages.

CONCLUSION

Taking RMB/USD exchange rate as the object of the empirical study, the exchange rate and the corresponding indicators are selected from July 2005 to January 2011 and the applied for the prediction respectively based on PCA-SVM, SVM and GA-BP. The compared predictions show that the prediction accuracy of PCA-SVM is the highest. And SVM model and the GA-BP models, this Study constructed PCA-SVM model achieved higher prediction accuracy. The MAPE value of PCA-SVM is 0.0038, less than that of SVM, indicating that PCA reduces the inputs and increases the prediction accuracy for SVM. The comparison between the PCA-SVM model and the GA-BP model comparison shows that PCA-SVM model has obvious advantages in a small number of samples. In addition, these three models all have higher prediction accuracy with MAPE below 1%, indicating that the proposed indicator system can accurately reflect the factors that affect the exchange rate.

The future work may lie in the study of the impact of non-economic indicators into the exchange rate, the expansion of the data samples and the efficient and stable method for optimizing the parameters for SVM.

ACKNOWLEDGEMENT

This study is funded by the Natural Science Foundation of Heilongjiang Province of China (G201013).

REFERENCES

Baghestani, H., 2009. Evaluating random walk forecasts of exchange rates. Stud. Econ. Finance, 26: 171-181.

Carriero, A., G. Kapetanios and M. Marcellino, 2009. Forecasting exchange rates with a large Bayesian VAR. Int. J. Forecast., 25: 400-417.

Cheng, M.Y., N.D. Hoang and Y.W. Wu, 2013. Hybrid intelligence approach based on LS-SVM and Differential Evolution for construction cost index estimation: A Taiwan case study. Automat. Constr., 35: 306-313.

Grossmann, A. and M.W. Simpson, 2010. Forecasting the Yen/U.S. dollar exchange rate: Empirical evidence from a capital enhanced relative PPP-based mode. J. Asian Econ., 21: 476-484.

Lan, X.J., L.X. Zhang, J.S. Lu and C.P. Chen, 2012. The mixed patterns recognition of control chart based on wavelet analysis and PSO-SVM. J. Zhejiang Univ. Technol., 40: 532-536.

Nelson, J.D.B., R.I. Damper and S.R. Gunn, 2008. Signal theory for SVM kernel design with applications to parameter estimation and sequence kernels. Neurocomputing, 72: 15-22.

Ranaee, V., A. Ebrahimzadeh and R. Ghaderi, 2010. Application of the PSO-SVM model for recognition of control chart patterns. ISA Trans., 49: 577-586.

Subasi, A. and M.I. Gursoy, 2010. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Syst. Appl., 37: 8659-8666.

Wu, C., Y.J. Guo and H. Xia, 2009. The model of credit risk assessment in commercial banks on fuzzy integral support vector machines ensemble. Oper. Res. Manage. Sci., 18: 115-119.