

2012 AASRI Conference on Computational Intelligence and Bioinformatics

Improved Support Vector Machine Oil Price Forecast Model Based on Genetic Algorithm Optimization Parameters

Xiaopeng Guo^{a,*}, DaCheng Li^a, Anhui Zhang^a

^a*School of Economics and Management, North China Electric Power University, Beijing 102206, China*

Abstract

An improved oil price forecast model that uses support vector machine (SVM) was developed. The new model, called the GA-SVM forecast model, is based on genetic algorithm (GA) optimization parameters. In traditional SVM models, penalty factor C and kernel function parameter σ are generally dependent on experience. These empirical parameters are difficult to accomplish the price data's change. Therefore, we used GA to optimize the parameter selection methods of SVM in accordance with training data, and improved SVM forecast precision. To verify the validity of the model, we selected and analyzed the Brent oil stock price data from 2001/12/27 to 2011/10/30. Data for 2009/07/30 to 2011/07/22 were designated as training data set, and those for 2011/08/08 to 2011/08/17 were employed for testing. Results show that the forecast efficiency of GA-SVM was better than that of traditional SVM.

2012 Published by Elsevier B.V. Selection and/or peer review under responsibility of American Applied Science Research Institute. Open access under [CC BY-NC-ND license](#).

Keywords: SVM, GA, Oil Price, Forecasting

1. Introduction

As one of the main energy sources, oil is widely applied to every field of the national economy. The prediction of oil price is an important issue related to oil production enterprises, oil consumption enterprises and national interests. Oil price is mainly influenced by international politics, economy, military affairs,

* Corresponding author. Tel.: +8613520328997.

E-mail address: guoxp2004@gmail.com.

diplomacy and other factors. The frequent change of these factors makes oil price show uncertainty, mutagenicity, randomness. Therefore, the accurate prediction of oil price has become a hotspot topic studied all over the world. Some researchers put forward to predicting oil price with hierarchy SVM model^[1], and present a forecasting model of oil price based on wavelet neural network and makes a simulation research^[2]. Prediction methods based on intelligent algorithm have also been widely applied in recent years, such as the prediction of coal price based on BP neural network^[3], a short-term electricity price forecasting model combined with wavelet transform, econometrics model and RBF neural network^[4], a time series forecasting model with variable structure based on RBF neural network^[5]. Good results have been obtained in the study above, but neural network, support vector machine model and some other models are inclined to fall into local extremum and overfitting. Therefore, it is a better solution to combine them with heuristic algorithm. From this viewpoint, this paper presents a SVM forecasting model of oil price selected by a genetic optimization (for short, GA) of the parameters.

2. Overview of SVM

SVM was put forward by Vapnik on the basis of small-sample statistical learning theory^[6], which is used primarily to study small samples under statistical learning rules, and is commonly adopted in pattern classification and nonlinear regression^[7,8].

The sample data set is given as $D = \{(x_i, y_i) | i = 1, 2, \dots, n\}$, where $x_i \in R^n$ represents the input variables and $y_i \in R$ denotes the output variables.

The SVM algorithm seeks one misalignment mapping from the input space to output space ϕ . Through this mapping, data X is mapped to a feature space Γ , and linear regression is carried out in the feature space with the following function:

$$\begin{aligned} f(x) &= [\omega \times \phi(x)] + b \\ \phi : R^n &\rightarrow \Gamma \end{aligned} \quad (1)$$

In (1), b is a threshold value. According to statistical learning theory, SVM determines the regression function through objective function minimization:

$$\begin{aligned} \min \left\{ \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i^* + \xi_i) \right\} \\ \text{s.t.} \quad y_i - \omega \cdot \phi(x) - b \leq \varepsilon + \xi_i^*; \\ \omega \cdot \phi(x) + b - y_i \leq \varepsilon + \xi_i; \\ \xi_i, \xi_i^* \geq 0. \end{aligned} \quad (2)$$

where C is a weight parameter for balancing the complex items of the model and training error, also called the penalty factor; ε is the insensitive loss function; and ξ_i^* and ξ_i are the relaxation factors. ξ_i^* is expressed as follows:

$$\xi_i^* = \begin{cases} 0, & |f(x) - y_i| \leq \varepsilon \\ |f(x) - y_i| - \varepsilon, & |f(x) - y_i| > \varepsilon \end{cases} \quad (3)$$

By solving the dual problem in (2), lag range factors a_i, a_i^* can be obtained, so that the regression equation coefficient is

$$\omega = \sum_{i=1}^n (a_i - a_i^*) x_i \quad (4)$$

The SVM regression equation is as follows:

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) K(X_i, X) + b \quad (5)$$

where $K(X_i, X)$ is the SVM kernel function. Kernel function types include linear kernels, polynomial kernels, and radial basis functions.

Penalty factor C , insensitive loss function ε , and kernel function parameter σ determine SVM performance. σ responds to the training data set characteristics, determines the complexity of the solution and affects the generalizability of the learning machine. Parameter C determines the penalty to large fitting deviation: An excessively large value may cause overlearning, but one too small easily results in less learning. The optimization of these parameters is therefore important in improving SVM performance.

3. SVM forecast model based on GA optimization parameters

Genetic Algorithm stems from the computer simulation research which carries on to the biological system. GA is an adaptive probability optimization technique devised by genetic and evolution mechanism of biology and is adequate for complex system optimization.

In genetic algorithm, the problem's solution is presented as the chromosome's survival of the fittest process. Through the operations of duplication, crossover and mutation etc. complete the chromosome group unceasing evolution, restrains finally in most adapts the environment individual, thus obtains the problem's optimal solution or the satisfactory solution.

The construction of the SVM forecast model based on GA optimization parameters (GA-SVM) entails 12 steps (Fig. 1).

- Step1. GA's binary coding.
- Step2. Selecting fitness function.
- Step3. Initializing GA's populations.
- Step4. Calculating fitness function, and judging if satisfying the convergence's requirement? ("Y", going to step 5. "N", going to step 8).
- Step5. Determine the optimal solution.
- Step6. Decoding the solution.
- Step7. Output parameters of SVM, c and g . The process is finished.
- Step8. Executing the duplicate operation.
- Step9. Judging if satisfying crossover probability. ("Y", going to step 10. "N", going to step11)

Step10. Executing crossover operation.

Step11. Judging if satisfying mutation probability. (“Y”, going back to step3, “N”, going to step12)

Step12. Executing mutation probability , and going back to step3.

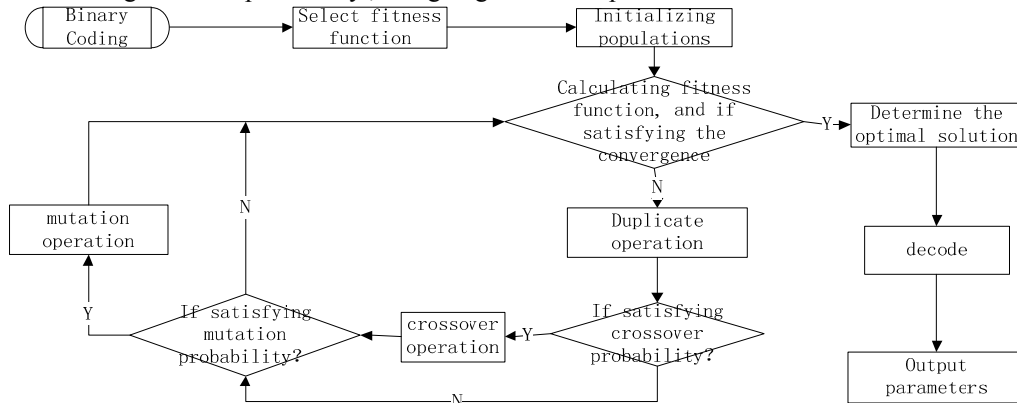


Fig. 1 GA-SVM modeling flow

4. Application case

This paper collected Brent oil stock price data from 2001/12/27 to 2011/10/30. The data between 2009/07/30 and 2011/07/22 were designated as training data set. On the basis of optimal SVM parameters by the training's, we constructed the SVM forecast model. This model was used to forecast the oil price of 2011/08/08~2011/08/17. To obtain better convergence results, we normalized the training data, while data testing and forecasting were carried out at a distribution between [0, 1]. The process was coded by Matlab and the LibSVM^[9] toolbox.

Predict results are shown in table 1, including original oil price data, GA-SVM-predicted data, traditional SVM-predicted data, and each predicted data set error and root mean squared error (RMSE). Errors and RMSE are important evaluation indicators of forecast results. The smaller the values obtained, the more accurate the model forecast.

Table 1. Original data and Predict data

Index	Original	ga-SVM		SVM	
		Predict	Error (%)	Predict	Error (%)
1	107.25	106.65	-0.55944	106.27	-0.91459
2	109.37	109.65	0.25601	109.54	0.15121
3	103.74	105.78	1.9665	105.85	2.0359
4	102.57	102.63	0.058497	103.3	0.70825
5	106.68	105.96	-0.67492	106	-0.63981
6	108.02	108.08	0.055545	108.05	0.027049
7	108.03	109.43	1.2959	109.63	1.4803
8	109.91	109.44	-0.42762	109.15	-0.69025
9	109.47	110.32	0.77647	110.58	1.0148

10	111.52	111.23	-0.26004	111.15	-0.33283
RMSE (%)	0.858		0.996		

As shown in Table 1, GA-SVM's RMSE is smaller than traditional SVM's RMSE. And mostly GA-SVM's errors are smaller than traditional SVM's errors. So GA-SVM is a more accurate forecast model than traditional SVM.

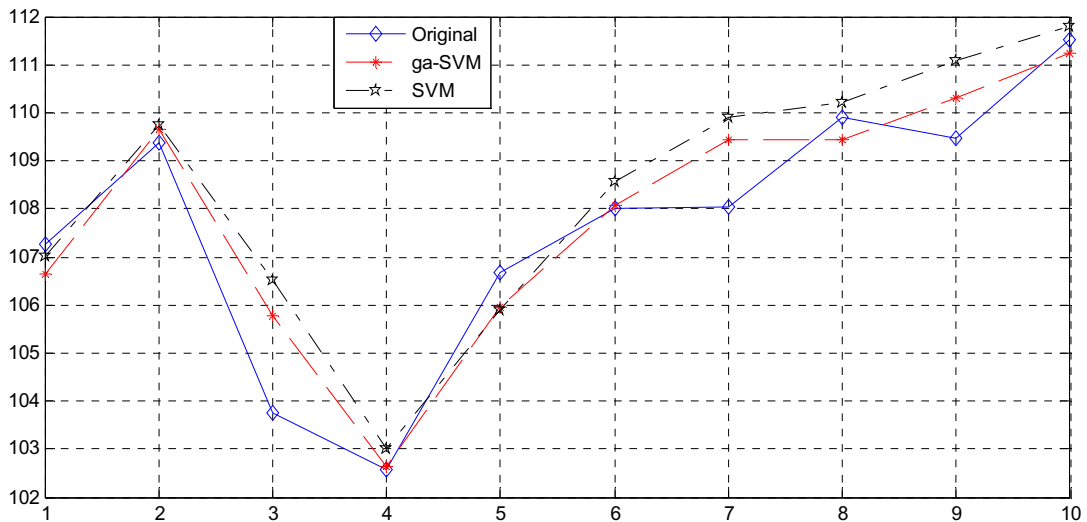


Fig.2 Comparison of the original and predict data

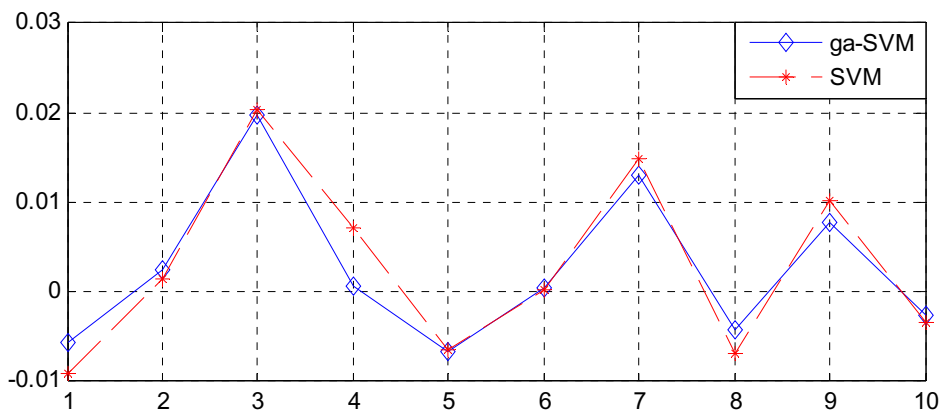


Fig.3 Comparison of the errors of the GA-SVM and traditional SVM

5. Summary

For traditional SVM, parameter selection algorithm always lead to some problems such as overlearning and underlearning, and diminish algorithm performance and affect forecast precision. Using the GA optimization choice, SVM penalty factor, and kernel function parameter yielded good results. Through original data confirmation and validating, the GA-SVM forecast model enables good forecast results.

In additional, GA's binary encoding method and the probability parameter's selection have a big influence to the algorithm performance. These questions are worth studying further.

Acknowledgements

Project supported by National Natural Science Fund of China (No. 71071054) and the Fundamental Research Funds for the Central Universities of China (No. 11QR34).

References

- [1] Hu Xue-mei, Zhao Guo-hao, Forecasting Model of Coal Demand Based on Matlab BP Neural Network[J]. Chinese Journal of Management Science, 2008,(16):521-525.
- [2] Zhang Dong-qing, Ma Hong-wei, Ning Xuan-xi, Time Series Prediction Based on Variable Structure RBF Neural Networks [J]. Chinese Journal of Management Science, 2010,(18):83-89.
- [3] Zhang Jin-liang, Tan Zhong-fu, Li Chun-jie, Short term electricity price forecasting of combined chaotic method[J]. Chinese Journal of Management Science, 2011,(19):133-139.
- [4] Zhu Xiao-mei, Guo Zhi-gang, Simulation Study on Forecasting Method of Oil Price Forecasting[J]. Computer Simulation, 2011,(28):361-364.
- [5] Wang Jun, Liu Zhi-bin, Oil Price Forecasting based on Hierarchical Support Vector Machine[J]. Computer Applications of Petroleum, 2009,(63):5-8.
- [6] Vapnik W N. The nature of statistical learning theory[M]. ZHANG Xue-gong, trans. Beijing: Tsinghua University Press, 2000.
- [7] Thissen U, Van Brakel R, de Weijer A P, et al. Using support vector machines for time series prediction[J]. Chemometrics and Intelligent Laboratory Systems, 2003, 69(2): 35–49.
- [8] Kim K J. Financial time series forecasting using support vector machines[J]. Neurocomputing, 2003, 55(1/2): 307–319.
- [9] Chih-Chung Chang, Chih-Jen Lin. LIBSVM : a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011,2(3):1-27.