

Application of SVR Models in Stock Index Forecast Based on Different Parameter Search Methods

Jiechao Chen^{1,2}, Huazhou Chen^{1*}, Yajuan Huo¹, Wanting Gao¹

¹College of Science, Guilin University of Technology, Guilin, China

²Bank of Hangzhou, Hangzhou, China

Email: *hzchengut@foxmail.com

How to cite this paper: Chen, J.C., Chen, H.Z., Huo, Y.J. and Gao, W.T. (2017) Application of SVR Models in Stock Index Forecast Based on Different Parameter Search Methods. *Open Journal of Statistics*, 7, 194-202.

<https://doi.org/10.4236/ojs.2017.72015>

Received: March 15, 2017

Accepted: April 17, 2017

Published: April 20, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Stock index forecast is regarded as a challenging task of financial time-series prediction. In this paper, the non-linear support vector regression (SVR) method was optimized for the application in stock index prediction. The parameters (C , σ) of SVR models were selected by three different methods of grid search (GRID), particle swarm optimization (PSO) and genetic algorithm (GA). The optimized parameters were used to predict the opening price of the test samples. The predictive results shown that the SVR model with GRID (GRID-SVR), the SVR model with PSO (PSO-SVR) and the SVR model with GA (GA-SVR) were capable to fully demonstrate the time-dependent trend of stock index and had the significant prediction accuracy. The minimum root mean square error (RMSE) of the GA-SVR model was 15.630, the minimum mean absolute percentage error (MAPE) equaled to 0.39% and the correspondent optimal parameters (C , σ) were identified as (45.422, 0.012). The appreciated modeling results provided theoretical and technical reference for investors to make a better trading strategy.

Keywords

CSI 300 Index, Support Vector Regression, Grid Search, Particle Swarm Optimization, Genetic Algorithm

1. Introduction

Stock index forecast is a non-linear dynamic system. There are many factors affecting the stock index, which goes with the complex fluctuation [1]. It had become a popular and interesting research issue to calculate the stock index to avoid the investment risk [2]. K-curve analysis was successfully applied to pre-

dict the trend of stock prices [3], but it couldn't accomplish the quantitative calculated. To quantitatively forecast the stock index price, traditional time-series models were introduced, such as autoregressive moving average model, which still failed in non-linear and non-stationary prediction [4]. At present, the research methods for stock index prediction vary from time series to artificial intelligence.

A variety of machine learning methods had been applied to stock index forecasting. They perform excellently with their merits of self-organization, self-learning and nonlinear approximation [5]. Support vector regression (SVR) is a kind of the simple, global optimization machine learning method, mainly described as nonlinear mapping transforming the low-dimensional data into a high-dimensional space, so that the data can be explained by a set of linear functions [6]. Improvement of SVR models depends on parameter optimization (regularization parameter C) and the selection of kernel functions [7] [8]. In this paper, the SVR models were trained by optimizing C and the kernel function. The radial basis function (RBF) was selected as the kernel. An over-large value of C possibly reduces the prediction ability of SVR models, and RBF kernel width (σ) commonly influences the model complexity [9] [10].

CSI 300 index is a capitalization-weighted stock market index designed to replicate the performance of 300 stocks traded in the Shanghai and Shenzhen stock exchanges. We established SVR calibration models to predict the opening price of CSI 300 index. In order to find out the optimizational modeling parameter combination of (C , σ), we tried to respectively utilize grid search (GRID) method, particle swarm optimization (PSO) and genetic algorithm (GA) for parameter selection. Furthermore, the opening price of CSI 300 index was predicted using the best parameters of the SVR model with GRID (GRID-SVR), the SVR model with PSO (PSO-SVR) and the SVR model with GA (GA-SVR). The rest of the paper is organized as follows. The methodology demonstrated in Section 1, Section 2 illustrated the SVR modeling process and prediction results. Section 3 contains the conclusions.

2. Experiments

2.1. Data Acquisition and Pretreatment

Daily trading data of CSI 300 index was scraped from the Wind Financial Terminal One-Stop Platform. The daily trading data (January 4, 2013 to November 30, 2016) were selected as sample set (a total of 949-day data). The data originally includes eight variables, which are opening price, ceiling price, the lowest price, closing price, charge rate, volume, turnover and the margin balance of China stock markets. We further calculated the 5-day average charge rate, the 20-day average charge rate, the 5-day average volume and the 20-day average volume as four new indicators to identify the index-changing trend. SVR models were established by the 949 samples with 12 variables to predict the opening price of the next day. The CSI 300 index daily opening price is shown in **Figure 1**. The sample set was normalized to the [0, 1] before establishing calibration models and the SVR prediction results were renormalized.

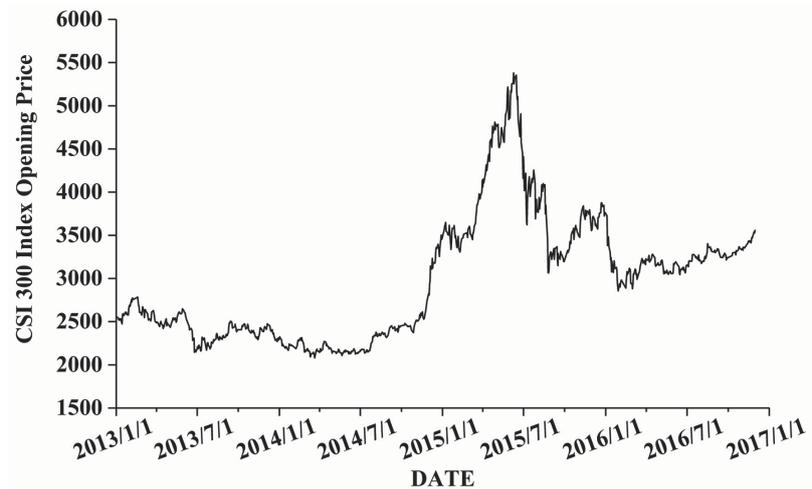


Figure 1. The CSI 300 index daily opening price.

2.2. Model Evaluation Indices

Four fifths of the total samples were selected for training, and rest one fifth for testing. The calibration models were practiced using the training samples and the parameters are optimized. Then, the training models with their parameters were applied to predict the opening price of the test samples. The prediction performance was evaluated using the root mean square error (RMSE) and mean absolute percentage error (MAPE). The formulas are as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n-1}}, \quad MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t|}$$

where y_t represents the real opening price, \hat{y}_t represents the predictive value of the opening price and n is total number of sample.

2.3. SVR Modeling and Parameter Search Method

Stock index prediction requires establishing an optimal prediction function based on stock history data and other interference to calculate the stock index price and to reveal the trend of the stock index. The function is defined as follows:

$$y_{t+1} = f(x_1, x_2, \dots, x_t),$$

where y_{t+1} represents the next-day opening price and x_1, x_2, \dots, x_t are input samples.

The SVR algorithm is used to estimate the function. The input data is mapped onto a high-dimensional feature space using RBF kernel. SVR is formulated as minimization of the following optimization problem,

$$\min_{\omega, \xi_i, \xi_i^*} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l [(\xi_i) + (\xi_i^*)],$$

where ω is the vector-form coefficients, ξ_i and ξ_i^* represent the relaxation factors. This optimization formulation can be transformed into the dual problem, and its solution is given by

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x_i, x^*) + b^*,$$

$$\text{St. } K(x, x^*) = \exp\left(-\|x - x^*\|^2 / 2\sigma^2\right),$$

where Lagrange multipliers (α_i, α_i^*) are controlled by C , $K(x, x^*)$ is a RBF kernel, σ represents the kernel width. C and σ are tuned respectively in the GRID-SVR, PSO-SVR and GA-SVR models respectively, to test the predictive capabilities.

The best parameter combination of (C, σ) is determined according to the minimum mean square error (MSE) under the 5-fold cross validation. The MSE is defined as follows,

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^l (y_t - \hat{y}_t)^2.$$

In the Grid search process, C and σ were pre-set in the tuning range of $\{2^{-8}, 2^{-7.5} \dots 2^{7.5}, 2^8\}$. They constitute a two-dimensional dynamic network. The optimal parameters (C, σ) are determined by searching the minimum MSE in this dynamic network.

In the PSO search process, a group of particles (random solutions of C and σ) were randomly initialized [11] [12]. And the optimal solution was found by iterations, with the training result (*i.e.* the minimum MSE) as the fitness value. In each iteration, the velocity and position of the particle swarm were globally and individually updated by searching the minimum MSE's. They were renewed by the following iterative equation,

$$v_d^i(t+1) = \omega \cdot v_d^i(t) + c_1 \cdot (p_{\text{best}}^i(t) - x_d^i(t)) + c_2 \cdot (g_{\text{best}}^i(t) - x_d^i(t)),$$

$$x_d^i(t+1) = x_d^i(t) + v_d^i(t+1),$$

where x_d represents the position of the particle, v_d represents the velocity of the particle, t is the number of iteration, i is the number of particles and ω is the velocity weight, c_1 and c_2 are learning factors. In this study, c_1 and c_2 were both valued 2, ω valued 0.5. The globally-optimal and individually-optimal velocity and position of the particle swarm were found by 100 iterative computations, so that the correspondent optimized parameters (C, σ) are determined.

In the GA search process, a population of chromosomes was randomly initialized [13] [14] [15]. And a new population was generated by selection, crossover, and mutation. The candidates were evolved toward better solutions with the MSE selected as the fitness value for iterative calculation. The evolution terminates when the maximum number of generations reached 100. The best SVR training model was obtained with the optimal parameters (C, σ) .

The GRID, PSO and GA methods were respectively applied for SVR parameters optimization. The flow charts of the experimental algorithms are shown in **Figure 2**. This figure depicted the entire modeling process.

3. Results and Discussion

The trading data of CSI 300 index from January 4, 2013 to November 30, 2016

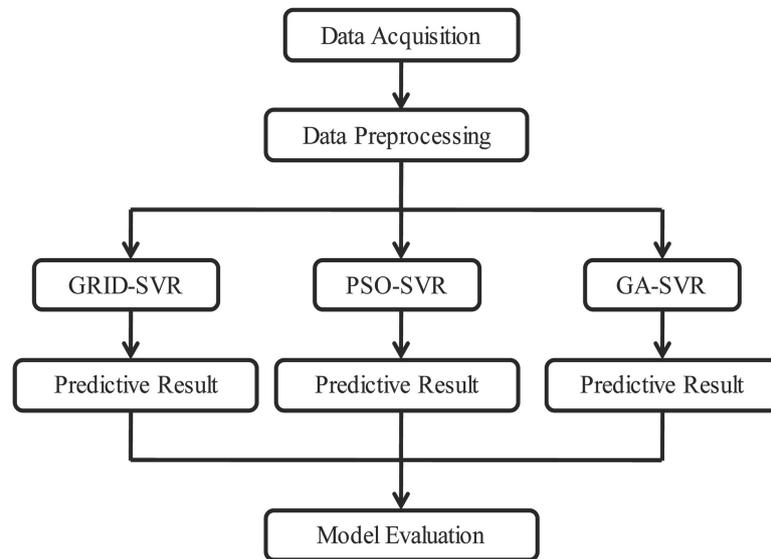


Figure 2. The flow charts of the experimental process.

was prepared for establishing the SVR calibration models. The records of previous consecutive 759 days were used as training samples and the remaining 190-day records as the test samples. Twelve variables of each sample were input to the SVR training models and a series of the next-day opening price were predicted. The parameters (C , σ) of SVR models were optimized respectively by GRID, PSO and GA.

During the GRID-SVR modeling process, the parameters (C , σ) were tuned for searching the minimum MSE. A larger value of C and a smaller value of σ generated a smaller MSE. The MSE contours are shown in **Figure 3**. The minimum MSE was found equaling to 7.967×10^{-5} , when C reaches 2^8 and σ reaches 2^{-7} (the solid point in **Figure 3**).

For the PSO-SVR models, an initial group of particles was randomly generated and then the positions and velocities of particles were globally and individually updated by 100 iterative computations. The MSE convergence process is shown in **Figure 4**. The MSE went rapidly to the minimum value (the solid point, equaling to 8.043×10^{-5}) at the 24th iteration, and it kept skipping reiteratively over the optimal value, the correspondent parameters (C , σ) are (60.576, 0.010).

In the GA-SVR parametric tuning, a group of candidate solutions was randomly initialized, and reiteratively evolved to a more appreciate alternative group of solutions by genetic selection, crossover, and mutation. **Figure 5** showed the MSE for each step of iteration. After 99 iterations, the minimum MSE found as 8.072×10^{-5} and the parameter (C , σ) are identified as (45.422, 0.012).

In summary, the optimally selected parameters (C , σ) of the SVR models were determined respectively for GRID, PSO and GA optimization. The GRID-SVR, PSO-SVR and GA-SVR calibration models with their corresponding optimal values of (C , σ) were applied to predict the validation samples. The parameters and prediction results are both presented in **Table 1**. As is shown in **Table 1**, the

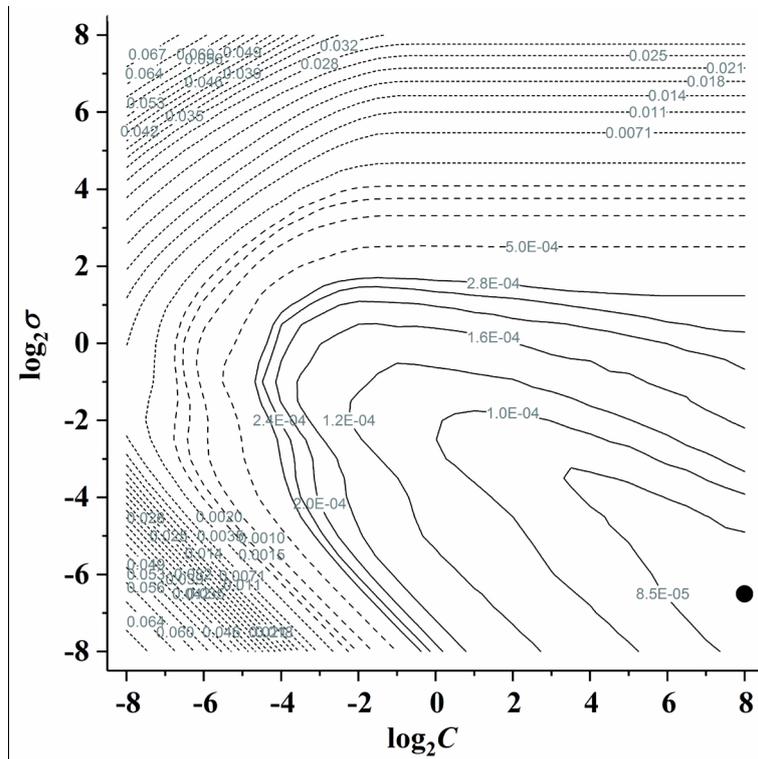


Figure 3. The GRID search of (C, σ) for SVR optimization.

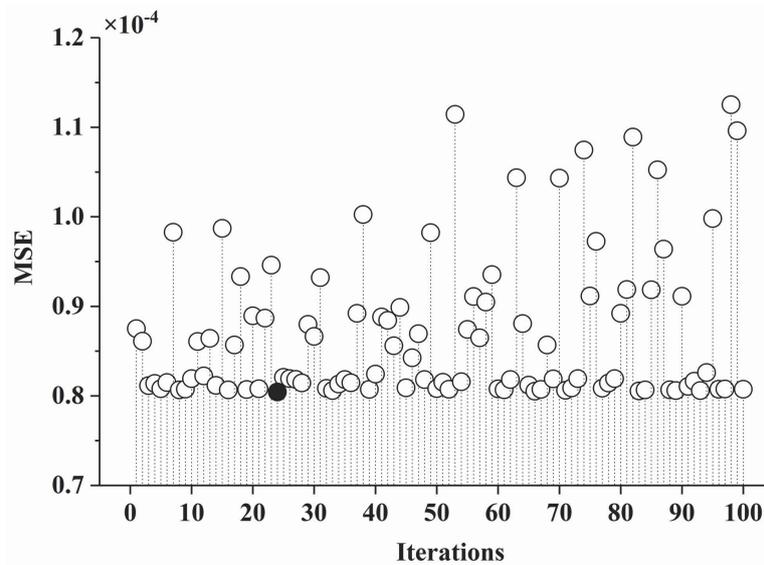


Figure 4. MSE of SVR optimization with PSO iteration.

obtained low values of RMSE and MAPE indicate that GRID, PSO and GA optimizing methods were acceptable for parametric optimization of SVR models, while the GA-SVR model was best validated. It provided the lowest RMSE of 15.630 and lowest MAPE of 0.39%. The comparison between the real and the GA-SVR predictive opening price is depicted in **Figure 6**. Therefore, the GRID-SVR, PSO-SVR and GA-SVR calibration models were feasible to accurately predict the short-term trend of opening price, and the GA-SVR had the highest prediction accuracy.

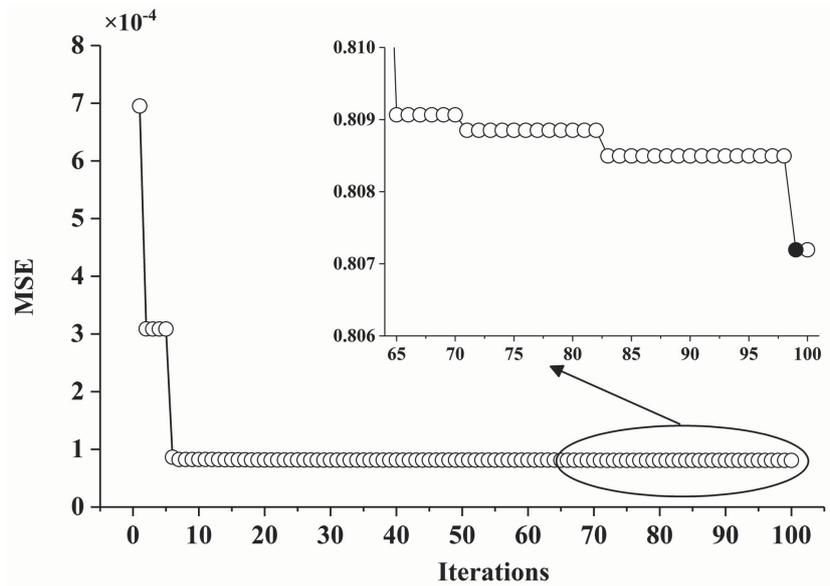


Figure 5. MSE of SVR optimization with GA evolution.

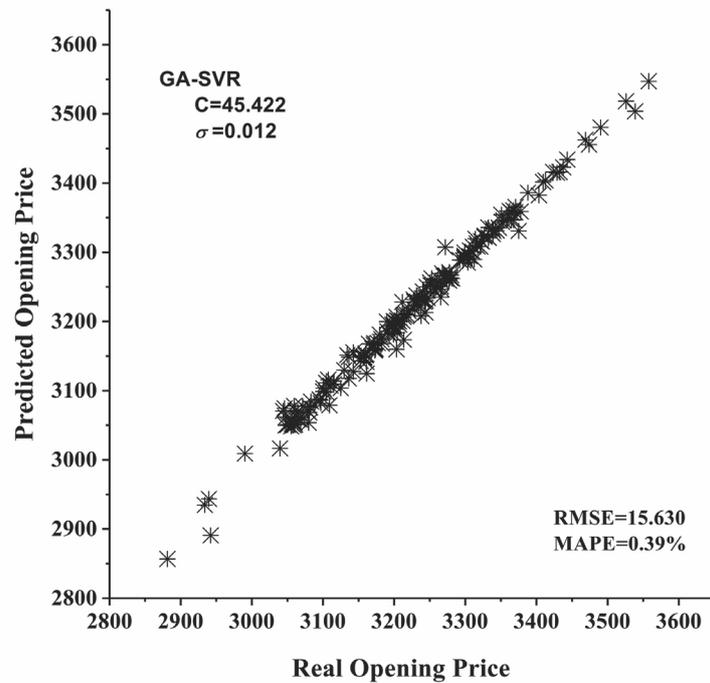


Figure 6. Comparison of the real opening price and the GA-SVR predicted opening price.

Table 1. Comparison of the predictive results for GRID-SVR, PSO-SVR and GA-SVR models.

Model Type	Model Parameters		Predictive Results	
	C	σ	RMSE	MAPE (%)
GRID Optimization	256	0.008	17.954	0.47
PSO Optimization	60.576	0.010	16.179	0.41
GA Optimization	45.422	0.012	15.630	0.39

4. Conclusion

In this study, the SVR models with GRID, PSO and GA parametric optimization were applied to predict the opening price of CSI 300 index. The optimal parameters (C , σ) were selected as (256, 0.008), (60.576, 0.010) and (45.422, 0.012) for GRID-SVR, PSO-SVR and GA-SVR models, respectively. The optimized SVR models were applied to the validation samples, obtaining the predictive RMSE's in the range of (15.63, 17.96), and the MAPE's ranged from 0.39% to 0.47%. The results showed that the GRID-SVR, PSO-SVR and GA-SVR calibration models were feasible to predict the short-term trend of opening price, and the GA-SVR had the most accurate prediction. The modeling performance provided theoretical and technical reference for investors to make a better trading strategy.

Acknowledgements

This work was supported by the National Natural Scientific Foundation of China (61505037), the Natural Scientific Foundation of Guangxi (2016GXNSFBA38-0077, 2015GXNSFBA139259).

References

- [1] Cao, Y., Liu, S. and Qiu, W. (2006) Research on Determinants of Intraday Price Movement in Shanghai Security Market. *System Engineering Theory and Practice*, **26**, 77-85.
- [2] Zheng, X. and Zhang, H. (2014) Application of Combination Forecasting Model in Stock Price Prediction. *Industrial Control Computer*, **27**, 121-122.
- [3] Wang, W., Yuan, Z., Xie, W. and Yang, J. (2009) Association Rule Study on Combination of Stock K Line. *Journal of Chengdu University (Natural Science Edition)*, **28**, 268-271.
- [4] Zheng, W. (2014) Short-Term Forecast of Stock Price of Shanghai Composite Index Based on ARIMA Model. *Economic Research Guide*, **234**, 136-137.
- [5] Tan, P., Steinbach, M. and Kumar, V. (2011) Introduction to Data Mining. Posts & Telecom Press, Beijing.
- [6] Vapnik, V.N. (1995) The Nature of Statistical Learning Theory. Springer Science + Business Media, New York. <https://doi.org/10.1007/978-1-4757-2440-0>
- [7] Gao, Z. and Yang, J. (2014) Financial Time Series Forecasting with Grouped Predictors Using Hierarchical Clustering and Support Vector Regression. *International Journal of Grid & Distributed Computing*, **7**, 53-64. <https://doi.org/10.14257/ijgdc.2014.7.5.05>
- [8] Cherkassky, V. and Ma, Y. (2004) Practical Selection of SVM Parameters and Noise Estimation for SVM Regression. *Neural Networks*, **17**, 113-126.
- [9] Zhu, Y. and Zhang, Y. (2003) The Study on Some Problems of Support Vector Classifier. *Computer Engineering and Applications*, **39**, 36-38.
- [10] Xiong, W. and Xu, B. (2006) Study on Optimization of SVR Parameters Selection Based on PSO. *Journal of System Simulation*, **18**, 2442-2445.
- [11] Chen, C., Tian, Y. and Bie, R. (2008) Research of SVR Optimized by PSO Compared with BP Network Trained by PSO. *Journal of Beijing Normal University (Natural Science)*, **44**, 449-453.

- [12] Gu, W., Chai, B. and Teng, Y. (2014) Research on Support Vector Machine Based on Particle Swarm Optimization. *Transactions of Beijing Institute of Technology*, **34**, 705-709.
- [13] Tang, K., Hu, G., Che, X. and Hu, L. (2010) Grid Host Load Prediction Model of Support Vector Regression Optimized by Genetic Algorithm. *Journal of Jilin University (Science Edition)*, **48**, 251-255.
- [14] Dai, H. (2008) Forecasting Population Based on Support Vector Regression with Intelligent Genetic Algorithms. *Computer Engineering and Applications*, **44**, 9-11.
- [15] Jiang, W., Wei, H., Qu, T. and Zhu, F. (2011) Predication of the Calorific Value for Fuel Coal Based on the Support Vector Regression Machine with Parameters Optimized by Genetic Algorithm. *Thermal Power Generation*, **40**, 14-19.



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ojs@scirp.org

