

Application of support vector regression algorithm in colleges recruiting students prediction

E ying

School of Foreign Languages
Harbin University of Science and Technology
Harbin 150040, China
eyingg@163.com

Abstract—Support vector regression algorithm is applied to colleges recruiting students prediction in the paper. As colleges recruiting students prediction is a nonlinear regression problem, the input training data of colleges recruiting students are nonlinearly mapped into a high dimensional space in support vector regression model. The amount of colleges recruiting students of Sichuan province from 2000 to 2008 is used to prove the effectiveness of support vector regression method. Then, the forecasting curves of support vector regression method and BP neural network and the comparison of forecasting error for amount of colleges recruiting students between support vector regression method and BP neural network are given in this study. The comparison results of forecasting error for amount of colleges recruiting students between support vector regression method and BP neural network indicate that support vector regression method has a higher forecasting accuracy than BP neural network.

Keywords—support vector regression algorithm; colleges recruiting students; forecasting performance

I. INTRODUCTION

In the past few years, artificial neural network is one of the most popular forecasting method. However, the shortcomings of artificial neural network lead to the limitations of practical application of artificial neural network[1,2]. Support vector regression algorithm is a new forecasting method[3], which is applied to colleges recruiting students prediction in the paper. As colleges recruiting students prediction is a nonlinear regression problem, the input training data of colleges recruiting students are nonlinearly mapped into a high dimensional space in support vector regression model[4-6]. The choice of the inappropriate parameters of support vector regression algorithm can lead to the bad prediction results of support vector regression algorithm. Thus, it is necessary to choose the appropriate parameters of support vector regression algorithm. In order to obtain the suitable parameters of support vector regression algorithm, genetic algorithm, based on Darwinian evolution principle[7-9], is applied to choose the appropriate parameters of support vector regression algorithm[7-9]. The amount of colleges recruiting students of Sichuan province from 2000 to 2008 is used to prove the effectiveness of support vector regression method. In this study, we give the forecasting curves of support vector regression method and BP neural network, and the comparison of forecasting error for amount of colleges recruiting students between support vector

regression method and BP neural network. The comparison results of forecasting error for amount of colleges recruiting students between support vector regression method and BP neural network indicate that support vector regression method has a higher forecasting accuracy than BP neural network.

II. SUPPORT VECTOR REGRESSION ALGORITHM

Colleges recruiting students prediction is a nonlinear regression problem. Support vector regression algorithm is a new forecasting method, which has a good nonlinear prediction performance.

The input training data of colleges recruiting students are nonlinearly mapped into a high dimensional space in support vector regression model.

Support vector regression function can be approximated as follows:

$$f(x) = w \cdot v(x) + b \quad (1)$$

where b is the bias and w is weight vector.

Define the regularized risk function to obtain the b and w .

$$U = \frac{1}{2} \|w\|^2 + C \frac{1}{n} \sum_{i=1}^n L_{\varepsilon}(y) \quad (2)$$

$$L_{\varepsilon}(y) = \begin{cases} |f(x) - y| - \varepsilon & |f(x) - y| \geq \varepsilon \\ 0 & |f(x) - y| < \varepsilon \end{cases}$$

where C is the regularization parameter.

Two positive slack variables are introduced to gain the following constrained form:

$$\text{Min} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right\} \quad (3)$$

Subject to

$$\begin{cases} y_i - (w, v(x_i)) - b \leq \varepsilon + \xi_i \\ (w, v(x_i)) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Then, the dual Lagrange form is obtained by using the Lagrange multipliers:

$$\text{Max} \left\{ \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle u(x_i), u(x_j) \rangle \right\} \quad (4)$$

Subject to

$$\begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases}$$

where α_i, α_i^* are the Lagrange multipliers.

Finally, the support vector regression function can be written as follows:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle v(x_i), v(x) \rangle + b \quad (5)$$

Define the kernel function $k(x_i, x)$,

$$k(x_i, x) = v(x_i) \cdot v(x) \quad (6)$$

Then, the support vector regression function can be written as follows:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (7)$$

The choice of the inappropriate parameters of support vector regression algorithm can lead to the bad prediction results of support vector regression algorithm. Thus, it is necessary to choose the appropriate parameters of support vector regression algorithm. In order to obtain the suitable parameters of support vector regression algorithm, genetic algorithm, based on Darwinian evolution principle, is applied to choose the appropriate parameters of support vector regression algorithm, which is shown in Fig.1.

Then, the steps of choosing the appropriate parameters of support vector regression algorithm by using genetic algorithm can be written as follows:

1) Code the parameters of support vector regression algorithm

Set the maximum iterative number to 100 and the size of the population of the chromosomes representing the parameters of support vector regression algorithm to 20.

2) Evaluate the fitness of each chromosome by fitness function F

Given the actual value t_i and the validation value t'_i , the fitness function F can be expressed as follows:

$$F(t_i, t'_i) = \left(1 + \frac{1}{l} \sum_{i=1}^l \left| \frac{t_i - t'_i}{t_i} \right| \right)^{-1} \quad (8)$$

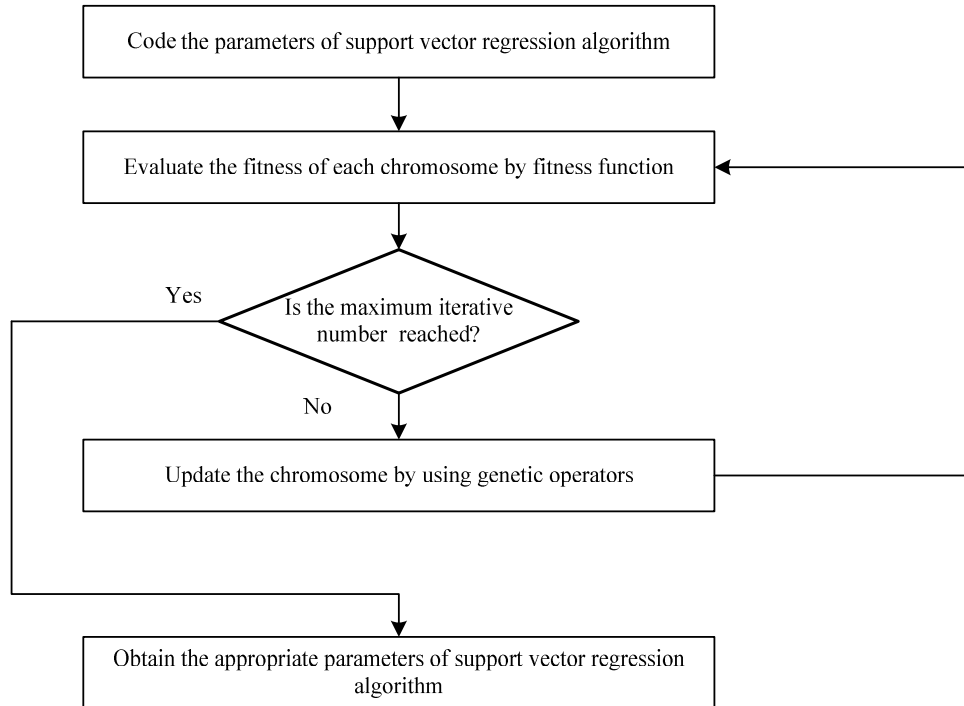


Figure 1. The steps of choosing the appropriate parameters of support vector regression algorithm

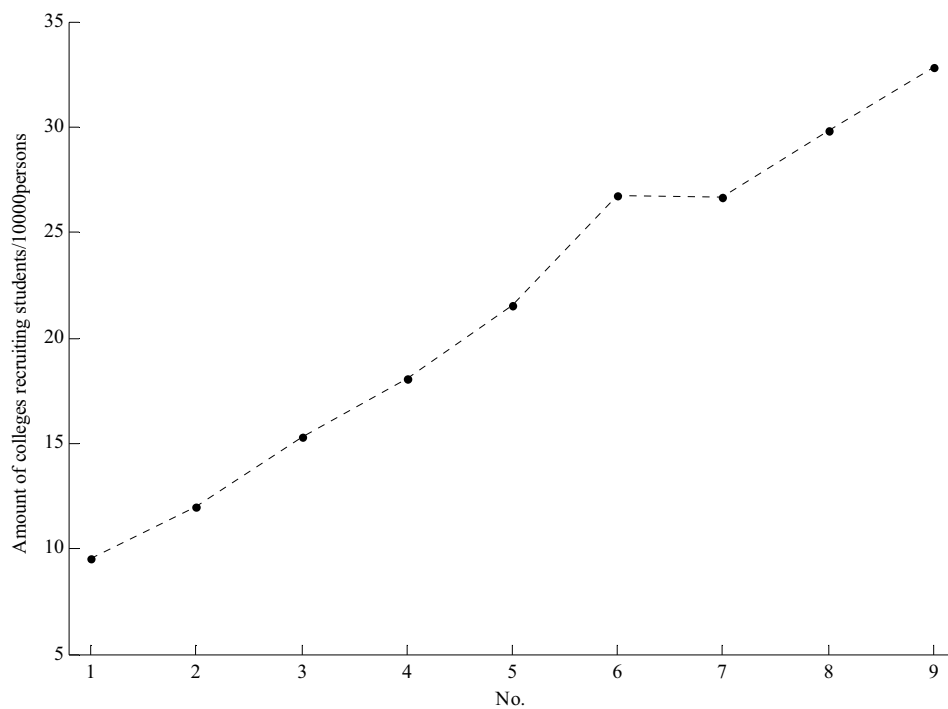


Figure 2. The amount of colleges recruiting students of Sichuan province from 2000 to 2008

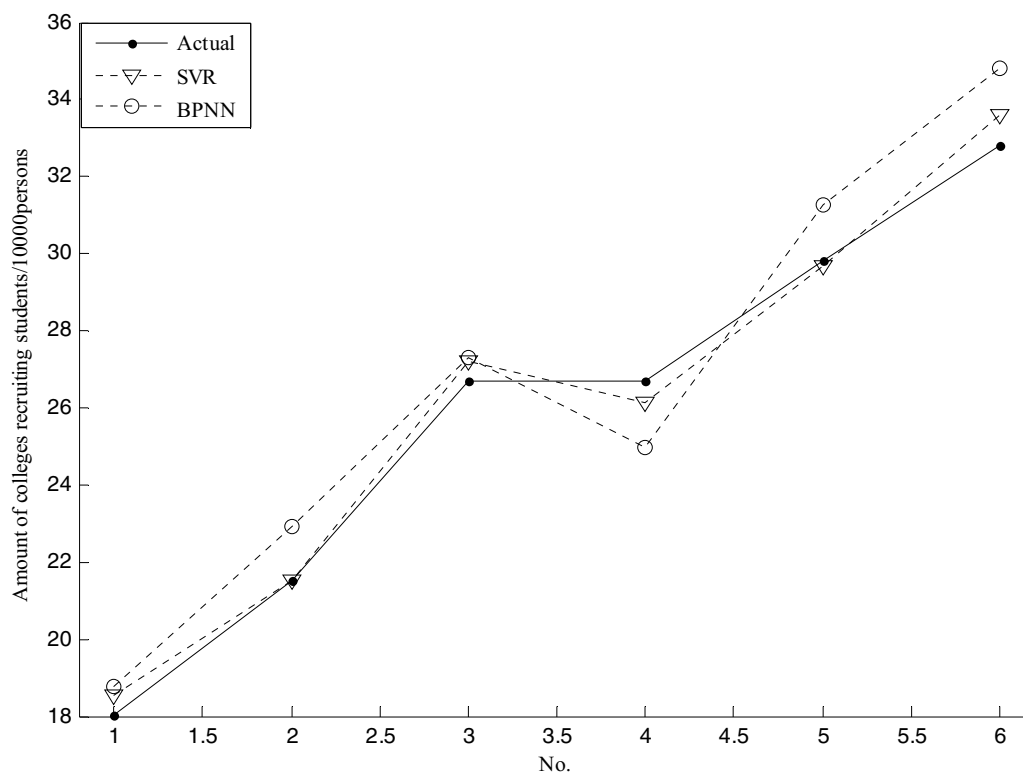


Figure 3. The forecasting curves of support vector regression method and BP neural network

3) Update the chromosome by using genetic operators

The new chromosome can be generated by using genetic operators including the selection, crossover, mutation operators.

4) Termination judgement for the computational process

The maximum iterative number is applied to termination judgement for the computational process.

III. EXPERIMENTAL TESTING AND ANALYSIS

Support vector regression algorithm is applied to colleges recruiting students prediction in this study. The amount of colleges recruiting students of Sichuan province from 2000 to 2008 is used to prove the effectiveness of support vector regression method.

Before the support vector regression model is trained, the training sample sets can be established as follows:

Define the input vector X

$$X = \begin{bmatrix} a_1 & a_2 & \cdots & a_t \\ a_2 & a_3 & \cdots & a_{t+1} \\ \vdots & \vdots & & \vdots \\ a_{n-t} & a_{n-t+1} & \cdots & a_{n-1} \end{bmatrix} \quad (9)$$

Define the corresponding output Y

$$Y = \begin{bmatrix} a_{t+1} \\ a_{t+2} \\ \vdots \\ a_n \end{bmatrix} \quad (10)$$

Then, support vector regression model is trained by the the training sample sets established by the above the input vector X and the corresponding output Y . In order to obtain the suitable parameters of support vector regression algorithm, genetic algorithm, based on Darwinian evolution principle, is applied to choose the appropriate parameters of support vector regression algorithm. In the genetic algorithm, we set the probability of crossover to 0.8, and set the probability of mutation to 0.01.

Fig.2 gives the amount of colleges recruiting students of Sichuan province from 2000 to 2008, where No.1~9 denote the years from 2000 to 2008. Fig.3 gives the forecasting curves of support vector regression method and BP neural network, where No.1~5 denote the years from 2003 to 2008, and we perform the comparison of forecasting error for amount of colleges recruiting students between support vector regression method and BP neural network. The comparison results of forecasting error for amount of colleges recruiting students between support vector regression method and BP neural network indicate that support vector regression method has a higher forecasting

accuracy than BP neural network.

IV. SUMMARY

Support vector regression algorithm is applied to colleges recruiting students prediction in the paper. In support vector regression model, the input training data of colleges recruiting students are nonlinearly mapped into a high dimensional space. As it is necessary to choose the appropriate parameters of support vector regression algorithm, genetic algorithm, based on Darwinian evolution principle, is applied to choose the appropriate parameters of support vector regression algorithm. The amount of colleges recruiting students of Sichuan province from 2000 to 2008 is used to prove the effectiveness of support vector regression method and we perform the comparison of forecasting error for amount of colleges recruiting students between support vector regression method and BP neural network. The comparison results of forecasting error for amount of colleges recruiting students between support vector regression method and BP neural network indicate that support vector regression method has a higher forecasting accuracy than BP neural network.

REFERENCES

- [1] George Kousiouris, Tommaso Cucinotta, Theodora Varvarigou, "The effects of scheduling, workload type and consolidation scenarios on virtual machine performance and their prediction through optimized artificial neural networks", *Journal of Systems and Software*, 2011, Vol.84, pp. 1270-1291.
- [2] Amir Hossein Alavi, Amir Hossein Gandomi, "Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing", *Computers & Structures*, 2011, Vol.89, pp.2176-2194.
- [3] Wei-Chiang Hong, "Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm", *Energy*, 2011, Vol.36, pp.5568-5578.
- [4] S.K. Lahiri, K.C. Ghanta, "Prediction of Pressure Drop of Slurry Flow in Pipeline by Hybrid Support Vector Regression and Genetic Algorithm Model", *Chinese Journal of Chemical Engineering*, 2008, Vol.16, pp.841-848.
- [5] S.G. Patil, S. Mandal, A.V. Hegde, "Genetic algorithm based support vector machine regression in predicting wave transmission of horizontally interlaced multi-layer moored floating pipe breakwater", *Advances in Engineering Software*, 2012, Vol.45, pp.203-212.
- [6] Senjian An, Wanquan Liu, Svetha Venkatesh, "Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression", *Pattern Recognition*, 2007, Vol.40, pp. 2154-2162.
- [7] Chung Hae Park, Woo Il Lee, Woo Suck Han, Alain Vautrin, "Improved genetic algorithm for multidisciplinary optimization of composite laminates", *Computers & Structures*, 2008, Vol.86, pp.1894-1903.
- [8] Sangjae Lee, Hyunchul Ahn, "The hybrid model of neural networks and genetic algorithms for the design of controls for internet-based systems for business-to-consumer electronic commerce", *Expert Systems with Applications*, 2011, Vol.38, pp.4326-4338.
- [9] Javier Puente, Alberto Gómez, Isabel Fernández, Paolo Priore, "Medical doctor rostering problem in a hospital emergency department by means of genetic algorithms", *Computers & Industrial Engineering*, 2009, Vol.56, pp.1232-1242.