TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.6, June 2014, pp. 4840 ~ 4848
DOI: 10.11591/telkomnika.v12i6.5843

Passenger Flow Forecasting using Support Vector Regression for Rail Transit

Bin Xia*[1], Fanyu Kong[2], Songyuan Xie[1]
[1]Logistic Engineering University, China.
[2]Chongqing Transport Planning Institute, China
*Corresponding author, email:xiabin126@126.com

Abstract

Support vector regression is a promising method for the forecast of passenger flow because it uses a risk function consisting of the empirical error and a regularized term which is based on the structural risk minimization principle. In this paper, the prediction model of urban rail transit passenger flow is constructed. It is to build an urban rail transit passenger flow forecast model and select the optimal parameters from the support vector regression through the variable metric method to obtain the minimal value from the LOO error bounds. The passenger flow is forecast by means of both support vector regression method and BP neural network method, and the results show that the support vector regression model has such theoretical superiority as minimized structural risk, thus having a higher forecasting accuracy under small sample conditions for short-term rail transit passenger flow, which predicts the promising forecasting performance that the method has.

Keywords*: rail transit, passenger flow, support vector regression, leave-one-bound*

1. Introduction

Rail transit passenger flow forecast is the foundation and quantitative basis for the planning, construction, operation and management of rail transportation, and the scientific accuracy of the results is directly related to the preparation of rail transportation planning and approval implementation of the project, it determines the development mode of rail transportation, road network scale, line alignments, hub settings and layout of interior space.

Currently, the rail transit forecast methods can be divided into three categories: 4-stage forecast method based on traffic demand analysis, disaggregate model forecast method, and forecast method not based on present passenger flow distribution.

The first forecast method is a conventional method commonly used both at home and abroad, which is achieved by collecting or using resident travel survey data, to split the transport modes in order to forecast the urban rail transit passenger flow on the basis of forecasting the total demand of urban passenger transport. It can affect the forecasting accuracy to a certain degree due to heavy workload of investigation, low data utilization, failure to take into account the reaction of traffic on land use. In view of the disadvantages of the 4-stage method, the disaggregate model was then raised, which is class I model in the unit of individuals that actually generate transport activities, to forecast the personal travel activities separately, and make statistics as per travel distribution, transportation modes and transit lines respectively, in order to get the total amount of traffic demand. Literature [1] uses the disaggregate model based on the 4-stage method to predict the passenger volume of urban rail transit. The above two methods are focused on the mid-/long-term forecast of passenger flow, but they cannot obtain effective forecast results in the case of dynamic changes in the recent passenger flow.

The third forecast method does not take into account the present passenger flow distribution, and usually, the forecast method is to transfer the present passenger flow of the relevant bus lines and bike traffic to the rail lines, so as to get a virtual base-year rail transit passenger flow; and it determines the growth rate of passenger rail transit passenger flow, and calculate the long-term rail transit passenger flow based upon the history data and growth law of relevant bus lines. Literature [1] applies gray theory to make a time series forecast on the annual urban rail passenger flow. Currently, most of these methods under research are to forecast and analyze the changes in passenger flow from the long-term perspective.

Received January 15, 2014; Revised March 18, 2014; Accepted March 30, 2014

The forecast results of short-term passenger flow to some extent determine the preparation and adjustment of transport organization plans and contingency plans for rail transit. If, in the case of dramatic changes in passenger flow during holidays and major events, a relatively accurate forecast can help provide effective decision support on the adjustment of the above two plans. Literature [2] applies the Fuzzy BP neural network model to make a data mining prediction on the scale of railway passenger flow, but because of the theoretical defects the neural network has, such methods have their final solution too much depended upon initial value and over learning, while having local minimality during the training process, relatively low rate of convergence, difficult choose of hidden network units, etc., so the forecasting results are not so easy to be promoted.

As a non-linear model forecasting method, support vector regression (SVR) has the following advantages compared with the neural networks, including global optimum always available during the training process, high generalization capability, solution space with sparesity, high rate of convergence and good forecasting performance under the small sample condition. Short-term passenger flow forecast is itself a complex random and non-linear process, and its variation has a high uncertainty. As a new algorithm based upon the structural risk minimization principle, SVR has a high accuracy and short time in forecasting the passenger flow under small sample conditions. This paper is based on the support vector regression algorithm, and builds forecasting models for short-term rail transit passenger flow by analyzing and mining history data and laws of passenger flow, and it provides a new way of thinking in carrying out transport organization, adjusting operational programs and preparing the contingency plan in a scientific and reasonable way.

2. Support Vector Regression Algorithm

Brief description of support vector regression algorithm [3] and set of the given training sample as:

$\{(x_{1,}y_1),\ldots,(x_k,y_k)\}$, and the optimal problem is:

$$\min_{w,\xi_i,\xi_i^*,b} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w}+\frac{C}{2}\sum_{i=1}^{k}\xi_i^{\,2}+\frac{C}{2}\sum_{i=1}^{k}(\xi_i^*)^2 \tag{1.a}$$

$$s.t. \quad -\xi_i^*-\varepsilon \le \mathbf{w}^T\phi(\mathbf{x}_i)+b-y_i \le \xi_i+\varepsilon \tag{1.b}$$

$$\xi_i,\xi_{\mathrm{i}}^* \ge 0\,,\quad i=1,\ldots,k \tag{1.c}$$

In Formula (1), the first item is aimed to maximize the classification interval, smoothing the function; the second and third items are error loss functions, for the purpose of reducing errors, constant $C > 0$, which is the degree of penalty ε beyond the error sample, ξ_i,ξ_i^*,ξ_i^* is the introduced slack variables. The sample input points are mapped by the function ϕ into a high dimensional space for linear regression ε is the insensitive loss function, and the error loss function uses the squares and forms of the minimum square function.

Using the duality principle, the Lagrange optimization method converts the above optimization problem into its dual problem:

$$\min_{\alpha_i,\alpha_i^*} \quad \frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^T\tilde{\boldsymbol{K}}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)+\sum_{i=1}^{k}(\alpha_i-\alpha_i^*)y_i+\varepsilon\sum_{i=1}^{k}(\alpha_i+\alpha_i^*)\,, \tag{2.a}$$

$$s.t. \quad \sum_{i=1}^{k}(\alpha_i-\alpha_i^*)=0\,,\quad 0\le\alpha_i,\alpha_i^*\,,\quad i=1,\ldots,k. \tag{2.b}$$

Where, $\alpha_i\alpha_i^*$ is the introduced Lagrange multiplier, α_i or α_i^* in the formula does not equal to zero, and the corresponding sample data is support vector.

Kernel Function $\boldsymbol{K}(\mathbf{x}_i,\mathbf{x}_j)=\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$, $\tilde{\boldsymbol{K}}=\boldsymbol{K}+\boldsymbol{I}/C$ in which $\boldsymbol{I}$ ▯ k dimensional unit matrix and the basic kernel functions include the following four types:

i) linear kernel function:

$$\boldsymbol{K}(\mathbf{x}_i,\mathbf{x}_j)=\mathbf{x}_i^T\mathbf{x}_j\ , \tag{3}$$

ii) Polynomial Kernel Function:

$$\boldsymbol{K}(\mathbf{x}_i\cdot\mathbf{x}_j)=(\gamma\mathbf{x}_i^T\mathbf{x}_j+r)^d,\gamma>0\ , \tag{4}$$

iii) Radial basis function (RBF):

$$\boldsymbol{K}(\mathbf{x}_i\cdot\mathbf{x}_j)=\exp(-\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2\Big/\sigma^2)\ , \tag{5}$$

iv) Sigmoid kernel function:

$$\boldsymbol{K}(\mathbf{x}_i,\mathbf{x}_j)=\tanh(\gamma\mathbf{x}_i^T\mathbf{x}_j+r)\ , \tag{6}$$

Literature [4] study shows: Linear kernel function cannot deal with the nonlinear input values; RBF kernel function has less high-dimensional kernel parameters than polynomial kernel functions, and in the SVM training process, the use of polynomial kernel function requires much more training time than that for the use of RBF kernel function; upon using Sigmoid kernel function, certain parameters have error values, so this paper adopts RBF as the input kernel function. *w,b* Which can be calculated with the following formula:

$$w=\sum_{i=1}^{k}(\alpha_i-\alpha_i^*)\phi(\mathbf{x}_i)\ , \tag{7}$$

$$b=y_l+\left(\tilde{\boldsymbol{K}}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)\right)_l-\varepsilon\ ,\ \text{where}\ \ \alpha_l>0 \tag{8}$$

$$b=y_l+\left(\tilde{\boldsymbol{K}}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)\right)_l+\varepsilon\ ,\ \text{wher}\ \ \alpha_l^*>0\ , \tag{9}$$

Obtain the regression function $f(\mathbf{x}_i)=w\cdot\phi(\mathbf{x}_i)+b$.

3. Study of the Selection of Model Parameters

Meanwhile, the forecasting performance of support vector regression is sensitive to the selection of parameters, but the choice of model parameters are mostly made through a simple cross-validation method (CV) or genetic algorithm (GA). Literature [4-6] Studies show that these two methods have the shortcomings of overly long training time during the selection process, CV can only use some of the samples for the calibration of parameters during the selection, while GA, in the selection process, cannot obtain the impact of a single parameter on the forecasting performance by the optimizing process, but in the use of Leave-one-out (LOO) for parameter selection, the optimal parameters can be obtained through seeking the minimization value from the least upper bound of the generalization error of the support vector machines. Compared with the above two methods, it has many advantages such as smaller time cost and many application parameters.

The support vector regression algorithm relies upon the choice of model parameters, and the adjustable model parameters of SVR are error penalty factor *C*, kernel function parameter σ^2 and insensitive loss function ε, where C and σ^2 are the independent variables for the above-mentioned kernel function $\tilde{\boldsymbol{K}}=\boldsymbol{K}+\boldsymbol{I}/C$, and all parameters are uniformly recorded as θ.

3.1. LOO Error Bound of Support Vector Regression

In this chapter, the minimum error bound method of LOO (Leave-one-out) is used for parameter selection and solution, and LOO error is a quantitative criteria used to characterize the degree of excellence of the support vector machine algorithm, which is defined as: selecting a sample in turn from *k* training samples as a test sample, and obtain the regression function with the remaining *k*-1 training samples as a training set, which are substituted into the input vector of the test sample to obtain the predicted value of the sample. Repeat *k* times to obtain the predictive value of each training sample, and finally to obtain the leave-one-out error value, with its expression as follows:

$$\mathrm{LOO}=\sum_{t=1}^{k}\left|f(x_t)-y_t\right| \tag{10}$$

In this formula, $f(x_t)$ and y_t are respectively the predicted value and the true value. When calculating the LOO error from the training set, it is necessary to use the algorithm *k* times for the training set that contains *k*-1 samples, to obtain *k* regression function values, so it is a heavy workload. Therefore, it is necessary to give up the accurate calculation of the LOO error value, and to use its upper bound value that is easily calculated instead. Meanwhile, such a property can be used that the LOO upper bound is the integral function of the parameter, θ, and the parameter value can be obtained by using the variable metric method for the LOO upper bound, so the parameter selection can be attributed to the optimization of seeking the LOO upper bounds.

Compared with the other LOO bounds, the radial interval upper bound of LOO has such advantages as simple operation and less promotion errors, etc., so the research adopts the LOO radial interval upper bound [3], with its expression as:

$$4\tilde{R}^2\mathbf{e}^T(\boldsymbol{\alpha}+\boldsymbol{\alpha}^*)+k\varepsilon \tag{11}$$

Where, $\mathbf{e}^T$ is the vector with the value of 1; $\tilde{R}$ is the minimum hyper-sphere radius that the input mapping function is contained in the high-dimensional space, with its expression as follows:

$$\tilde{R}=\min_{c}\left\{\tilde{R}\,\middle|\,\left\|\tilde{\phi}(\mathbf{x}_i)-\boldsymbol{c}\right\|\le\tilde{R}, i=1,\dots,k\right\} \tag{12}$$

Here, *c* is the center of the hypersphere.The improved input mapping function is defined as:

$$\tilde{\phi}(\mathbf{x}_i)=\begin{bmatrix}\phi(\mathbf{x}_i) & \dfrac{e_i}{\sqrt{C}}\end{bmatrix}^T \tag{13}$$

Where, e_i is the unit vector, $\phi(\mathbf{x}_i)$ is the input mapping function in Formula (1).

3.2. Gradient Calculation

When obtaining the optimal parameter of the model by using the DFP method, the gradient of $\tilde{R}^2$ and $\boldsymbol{\alpha}+\boldsymbol{\alpha}^*$ in Formula (11).

$\tilde{R}^2$ is the optimal value [13] of the following issues:

$$\max_{\lambda}\ \sum_{i=1}^{k}\lambda_i\tilde{K}_\theta(\mathbf{x}_i,\mathbf{x}_i)-\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\tilde{K}_\theta(\mathbf{x}_i,\mathbf{x}_j) \tag{13a}$$

$$\sum_{i=1}^{k}\lambda_i=1,\ \ 0\le\lambda_i,\ \ i=1,\dots,k \tag{13b}$$

So the gradient of $\tilde{R}^2$ can be expressed as:

$$\frac{\partial\tilde{R}^2}{\partial\theta}=\sum_{i=1}^{k}\lambda_i\frac{\partial\tilde{K}_\theta(\mathbf{x}_i,\mathbf{x}_i)}{\partial\theta}-\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\frac{\partial\tilde{K}_\theta(\mathbf{x}_i,\mathbf{x}_j)}{\partial\theta} \tag{14}$$

The following is the derivation of $\partial\boldsymbol{\alpha}+\boldsymbol{\alpha}^{*}/\partial\theta$. For the optimal solution that meets the dual problem (2), Formula (9) and Formula (11) are established and can be written in the matrix form:

$$\mathbf{G}\begin{bmatrix}\hat{\boldsymbol{\alpha}}\\ b\end{bmatrix}=\begin{bmatrix}\mathbf{p}\\ 0\end{bmatrix} \tag{15}$$

Where, $\mathbf{G}=\begin{bmatrix}\tilde{\boldsymbol{K}} & \mathbf{e}\\ \mathbf{e}^{\mathrm{T}} & \mathbf{0}\end{bmatrix}$, $\hat{\boldsymbol{\alpha}}=\boldsymbol{\alpha}^{*}-\boldsymbol{\alpha}$, $p_i=\begin{cases}y_i-\varepsilon, & \hat{\alpha}_i>0\\ y_i+\varepsilon, & \hat{\alpha}_i<0\end{cases}$. It is easy to know that the following formula is established.

$$\frac{\partial(\alpha_i+\alpha_i^{*})}{\partial\theta}=\delta_i\frac{\partial\hat{\alpha}_i}{\partial\theta} \tag{16}$$

Where $\delta_i=\begin{cases}1, & \hat{\alpha}_i>0\\ -1, & \hat{\alpha}_i>0\end{cases}$, for the parameter ε except $\theta\left(C,\sigma^{2}\right)$, by differentiating from Formula (15), the following formula is established:

$$\mathbf{G}\begin{bmatrix}\dfrac{\partial\hat{\alpha}}{\partial\theta} & \dfrac{\partial b}{\partial\theta}\end{bmatrix}^{T}+\frac{\partial\mathbf{G}}{\partial\theta}\begin{bmatrix}\hat{\alpha}\\ b\end{bmatrix}=0 \tag{17}$$

The above formula is deformed, and the following formula can be obtained:

$$\frac{\partial\hat{\alpha}}{\partial\theta}=\mathbf{G}^{-1}\left(\frac{\partial\tilde{K}_{\theta}}{\partial\theta}\hat{\alpha}\right) \tag{18}$$

When the parameter θ is ε, there is the following equation that holds:

$$\frac{\partial\hat{\alpha}}{\partial\varepsilon}=\mathbf{G}^{-1}\frac{\partial\mathbf{p}}{\partial\varepsilon} \tag{19}$$

3.3. DFP Algorithm

Let f be the function of the desired parameter θ^{k}, and here, the BFGS variable metric algorithm is used to solve the parameter. Let θ^{k} be the parameter of the k-th iteration, and $f(\theta^{k})$ is the objective function, the algorithm is as follows:

a. Calculate the search direction, $p=-H_k\nabla f(\theta^{k})$.

b. Make one-dimensional search in the p-direction, and determine the optimum step size λ based on the following formula:

$$\min_{\lambda} f(\theta^{k}+\lambda p) \tag{20}$$

From the above formula, the next parameter point $\theta^{k+1}=\theta^{k}+\lambda p$ can be obtained.

c. Calculation

$$H_{k+1}=\begin{cases}\left(I-\dfrac{sy^{T}}{y^{T}s}\right)H_k\left(I-\dfrac{ys^{T}}{y^{T}s}\right)+\dfrac{ss^{T}}{y^{T}s} & if \quad y^{T}s>0\\ H_k, & \end{cases} \tag{21}$$

Where, $s=\theta^{k+1}-\theta^{k}$, $y=\nabla f(\theta^{k+1})-\nabla f(\theta^{k})$, H_k is the unit matrix taken from the first iteration.

d. The iteration termination condition is as follows:

$$\frac{\left[f(\theta^{k-1})-f(\theta^{k})\right]}{f(\theta^{k-1})}<\varepsilon \tag{22}$$

In the iteration process, seek the optimal solution from the logarithm value θ of the parameter $\left(\ln C,\ln\sigma^{2},\ln\varepsilon\right)$, and in the iteration algorithm, the gradient is $\frac{\partial f}{\partial\ln\theta}=\theta\frac{\partial f}{\partial\theta}$.

4. Instances of Urban Rail Transit Passenger Flow Forecast

4.1. Model Construction

The passenger flow forecast of short-term urban rail transit is based on the dynamic changes in historical passenger flow data, to determine the historical data that impact the future passenger flow as the input dimension of the model, and the predictive value as the output dimension of the model through analysis and arrangement of the time series data of passenger flow. If there are n historical data for the selected model, then the model is constructed to make regression forecasts on the n+1 dimensional hyperplane. The specific forecasting model can be expressed as:

$$v_i(t+1)=b_1v_i(t)+b_2v_i(t-1)+...+b_nv_i(t-n+1) \tag{23}$$

Where, b_j is the weighting coefficient, $v_i(t+1)$ is the rail transit passenger flow for the $t+1$ period. The parameters (C,σ^2,ε) in the support vector regression model has a significant impact on the forecasting performance of the model, and the selection of parameters is usually made with the Cross Validation method, that is, to divide the forecasting sample values into m groups, and select a group of parameters and train the m-1 group of data from it, and the rest one serves as the checksum value of forecasting performance under that parameter. After several trainings and verifications, an optimal set of parameters can be determined to forecast the future value.

4.2. Forecasting Results and Analysis

This paper is focused on the support vector regression algorithm, using LIBSVM software package to forecast the historical data of passenger flow on a certain route of Shanghai subway on a daily basis, where the training sample set is the historical data for the first five weeks, the forecasting sample set is the 7 passenger values for the sixth weeks, and the number of cross-validation groups, m, is 5, at the same time, choose the 3-layer BP neural network method for controlled trial, and use the average relative error, root mean square error, maximum relative error, and minimum relative error as the evaluation indexes.

With the cross-validation process, the support vector regression parameters can be ultimately determined; BP neural network parameters choose them as: input layer 7, intermediate layer 13, output layer 7; Learning Network Operators as 0.8, damping coefficient as 0.1, error adjustment factor as 0.2, error objective function as 0.05, and the regression forecasting results is shown in Table 1 $(C,\sigma^2,\varepsilon)=(71.1,5.4,0.063)$.

Table 1. Forecasting Sample Result Values

No.	Actual value/person	BP forecast value/person	SVR forecast value/person
36	599956	521313	555779
37	549094	525005	535026
38	557770	535809	536650
39	545399	543833	533642
40	604255	640660	601646
41	594073	640129	563103
42	564897	556048	544003

Figure 1 shows the relative error absolute value function curve for the forecasting regression results in both ways. As seen from the figure, except for a few points, the relative error absolute values for the support vector regression (SVR) are mostly smaller than the error values of the BP neural network. Table 2 compares the statistical results from forecasting the relative error absolute values under the two methods, and the statistical results in this table show that except for the minimum relative errors, the rest of the statistical indicators are lower than the latter, and their forecasting results are relatively stable, with less fluctuation, which indicates that the forecasting performance of SVR is better than the BP neural network.

Table 2. Statistics of Relative Error Forecasting

Indicators	BP	SVR
Average relative error (%)	5.2957	2.8052
Root mean square error(%)	6.9827	2.0292
Maximum relative error (%)	7.75	5.3633
Minimum relative error (%)	0.229	0.4318

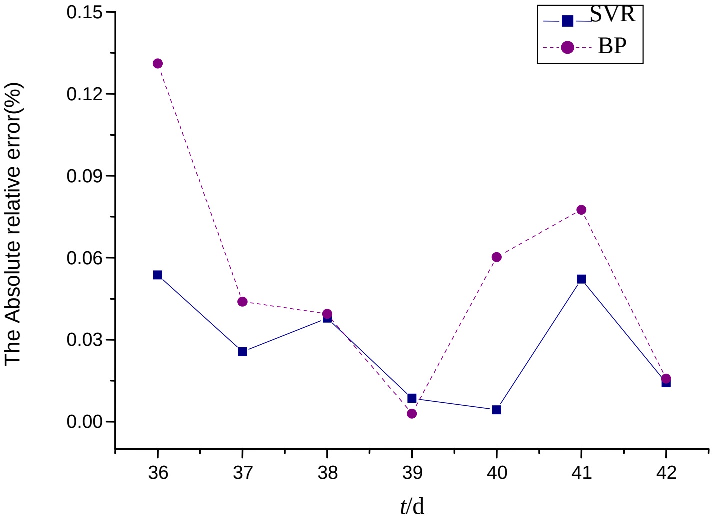

Figure 1. The Absolute Relative Errors of Forecast Samples as a Function of t

4.3. Parameter Discussion

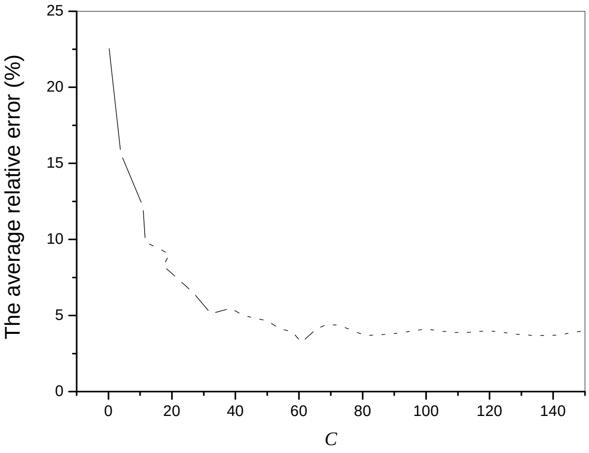

Figure 2. The Relationship between the Average Relative Error for Forecasting Samples and the Parameter *C*

Based upon the parameters obtained from the optimization LOO upper bound process, make an analysis of the impact of the changes in a single parameter on the forecasting performance. Choose the optimal parameter $\left(\ln C, \ln \sigma^2, \ln \varepsilon\right)$ = (3.8, 0.92, -2.13), that is, $(C, \sigma^2, \varepsilon)$ = (44.7, 2.50, 0.12) serves as the fixed condition for the study of single parameters.

As can be seen from Figure 2, when σ^2 and ε are fixed, the average relative error of the forecasting sample decreases with the increase of *C*, and it comes to a stable status when increasing to the vicinity of the optimal parameter; as seen from Figure 3, when ε increases, the average relative error of the forecasting sample has a relatively stable change when it is less than 0.154, but it begins to increase rapidly when it is greater than 0.154; in Figure 4, with the increase of σ^2 ,the average relative error of the forecasting sample is in a relatively stable change. As can be seen from the accuracy change curve of the forecasting sample, the three parameters are in the vicinity of the optimum value, with a large allowable range and a relatively small change in accuracy, which is similar to the research results in Literature [3].

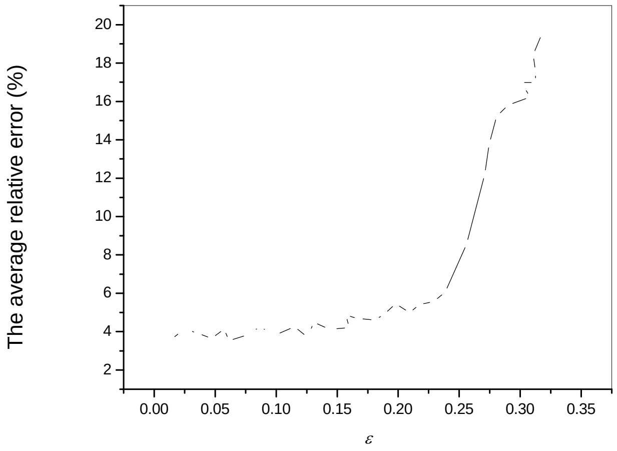

Figure 3. The Relationship with the Average Relative Error of the Forecasting Sample with the Parameter ε

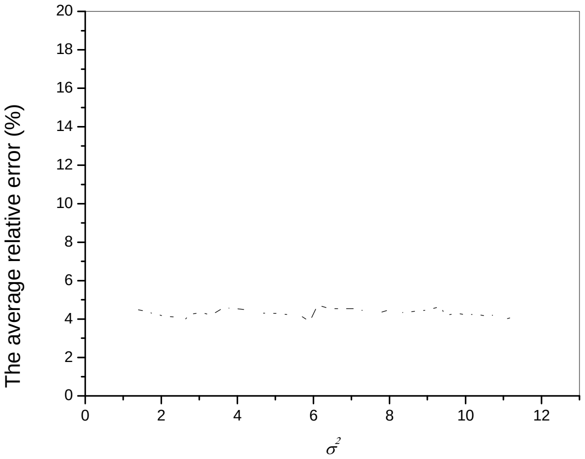

Figure 4. The Relationship with the Average Relative Error of the Forecasting Sample with the Parameter σ^2

5. Conclusion

This paper analyzes and discusses several models of the urban rail transit passenger flow forecasting, compares the range of applications and research status of various models, and proposes the necessity of forecasting the short-term rail transit passenger flow. It forecasts short-term passenger flow by means of both support vector regression method and BP neural network method, and the results show that the support vector regression model has such a theoretical superiority as minimized structural risk, thus having a higher forecasting accuracy under small sample conditions for short-term urban rail transit passenger flow, which predicts a promising forecasting performance the method has.

References

[1] Wu Qiang. The Application of Grey Forecasting Method in urban rail transit passenger flow forecasting. *Study on Urban Rail Transit.* 2004; 7(3): 52-55.

[2] Wang Yanhui. Railway Passenger Traffic Volume Data Mining Forecasting Method and Its Application. *Journal of Railway.* 2004; 26(5): 1-7.

[3] Ming-Wei Chang, Chih-Jen Lin. Leave-one-out Bounds for Support Vector Regression Model Selection. *Neural Computation*. 2005; 17(5): 1188-1222.

[4] Kyoung jae Kim. Financial time series forecasting using support vector machines. *Neurocomputing*. 2003; 55 (3): 307-319.

[5] Friedrichs F, Christian I. Evolutionary tuning of multiple SVM parameters. *Neurocomputing*. 2005; 64: 107-117

[6] Chapelle O, Vapnik V. Choosing multiple parameters for support vector machines. *Machine learning*. 2002; 46: 131-159.