Road Traffic Freight Volume Forecast Using Support Vector Machine Combining Forecasting

Shang Gao
School of Computer Science and Technology, Jiangsu University of Science and Technology, Zhenjiang 212003, China
Email: gao_shang@hotmail.com

Zaiyue Zhang and Cungen Cao
Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
Email: yzzjzzy@sina.com, cgcao@ict.ac.cn

***Abstract*—The grey system forecasting model, neural network forecasting model and support vector machine forecasting model are proposed in this paper. Taking the road goods traffic volume from year of 1996 to 2003 in the whole country as a study case, the forecasting results are got by three methods. From the forecasting results, we can conclude that the accuracy of the support vector machine forecasting method is higher. Analyzing the characteristic of combining forecasting method, based on grey system forecasting model, neural network forecasting model and support vector machine forecasting model, the linear combining forecasting model, neural network combining forecasting model and support vector machine combining forecasting model are set up. Compared with single prediction methods, linear combining forecasting method and neural network combining forecasting model, the accuracy of the support vector machine combining forecasting method is higher.**

***Index Terms*—grey system, neural network, support vector machine, combining forecasting, traffic volume**

I. Introduction

With the development of society, transportation plays a more and more important role in social economical progress, at the same time transportation is rapidly progressing too. As far as we know, the accurate and objective prediction of the future road transportation demand is the just foundation of the scientific transportation planning. It turns out that the excepted effect gains verification, effectively improves the model accuracy, and makes more exact forecasting, which expects to be helpful to concerned departments and personnel for them to grasp the traffic market trend or make decision. Combining forecasting has brought great attention to by the forecasting circles since 1969 when J . M. Bates and C. W. J. Granger proposed its theory and method. The theory and methods of combining forecasting have been developed widely in recent years [1]. For practical cases of various forecasting problems, combining forecasting models may have different forms. Among them proportional mean combining forecasting models are widely used , such as simple weighted arithmetic proportional mean combining forecasting model , simple weighted square root proportional mean combining forecasting model , simple weighted harmonic proportional mean combining forecasting model , generalized weighted arithmetic proportional mean combining forecasting model and generalized weighted logarithmic proportional mean combining forecasting model , etc. In this paper, the grey system forecasting model, neural network forecasting model and support vector machine forecasting model are proposed. Based on grey system forecasting model, neural network forecasting model and support vector machine forecasting model, the linear combining forecasting model, combining support vector machine forecasting model are set up.

II. Three Predicting Methods

A. Grey Predicting Model

Since 1982, there has been a quick development in grey systems theory in China, and it is also very successful in the application of the theory to many real projects, such as agriculture, society, economics, engineering, IT, data mining, management, biological protection, robot, ecology, image processing, environmental studies, etc.. Grey model GM(1,1) due to whole distinguishing features: modeling by less data (suiting the data as few as 4), thus underlay grey modeling and grey forecasting. Because sometimes the precision of grey method by means of AGO (accumulated generation operation) and IAGO (inverse accumulated generation operation) can not meet the requirement of actual forecasting, much research in theory and application has been done.

The GM (1,1) model means a single differential equation model with a single variation. The modeling process is as follows: First of all, observed data are converted into new data series by a preliminary transformation called AGO (accumulated generating operation). Then a GM model based on the generated sequence is built, and then the prediction values are obtained by returning an AGO' s level to the original

doi:10.4304/jsw.6.9.1680-1687

level using IAGO (inverse accumulated generating operation).

Now we introduce the grey predicting model GM(1,1). Let $X^{(0)}=\{x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\}$.By defining

$$x^{(1)}(k)=\sum_{i=0}^{k}x^{(0)}(i),$$

We get a new series $X^{(1)}=\{x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\}$.

To some processes, $X^{(1)}$ is the solution of the following grey ordinary differential equation [2]

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=b \tag{1}$$

where a and b are grey numbers. The equation (1) is called GM (1, 1).

By taking average $z^{(1)}(k)=\frac{1}{2}[x^{(1)}(k)+x^{(1)}(k-1)]$ and other transformations, we get that

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \\ \\ x^{(0)}(n) \end{bmatrix}=\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} \tag{2}$$

which can be simplified as $y_N=B\hat{a}$, where $y_N=[x^{(0)}(2),x^{(0)}(3),\cdots,x^{(0)}(n)]^T$, $\hat{a}=\begin{bmatrix} a & b \end{bmatrix}^T$ and $B=\begin{pmatrix} -z^{(1)}(2) & -z^{(1)}(3) & \cdots & -z^{(1)}(n) \\ 1 & 1 & \cdots & 1 \end{pmatrix}^T$.

If rank(B)=2, the equation (2) has a unique solution: $\hat{a}=(B^TB)^{-1}B^Ty_N$. Therefore, from (1) we obtain the generating model:

$$\hat{x}^{(1)}(k+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-ak}+\frac{b}{a} \tag{3}$$

From (3), each value of $\hat{x}^{(0)}(k)$ can be computed. Thus, we compute the feedback values $\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)$:

$$\begin{cases} \hat{x}^{(1)}(1)=x^{(0)}(1) \\ \hat{x}^{(0)}(k)=(x^{(0)}(1)-\frac{b}{a})(1-e^{a})e^{-a(k-1)} \qquad (k=2,3,\cdots) \end{cases} \tag{4}$$

The road goods traffic volumes from year of 1996 to 2003 in the whole country are listed in Table 1 in detail. We can get the forecasting values of grey GM(1,1) listed in Table 1.

TABLE I.
THE ACTUAL TRAFFIC VOLUME FROM 1996 TO 2003 AND THE FORECASTING VALUES OF GM

Time	Actual Data	forecast value	relative error
1996	984	984.0	0.00%
1997	977	950.2	2.75%
1998	976	980.1	0.42%
1999	990	1010.9	2.11%
2000	1039	1042.7	0.36%
2001	1056	1075.5	1.85%
2002	1116	1109.4	0.59%
2003	1160	1144.3	1.36%
2004		1180.3	
2005		1217.4	
2006		1255.7	
2007		1295.2	
2008		1336.0	

B. Neural Network Predicting Model

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons.

Many important advances have been boosted by the use of inexpensive computer emulations. Following an initial period of enthusiasm, the field survived a period of frustration and disrepute. During this period when funding and professional support was minimal, important advances were made by relatively few researchers. These pioneers were able to develop convincing technology which surpassed the limitations identified by Minsky and Papert. Minsky and Papert, published a book (in 1969) in which they summed up a general feeling of frustration (against neural networks) among researchers, and was thus accepted by most without further analysis. Currently, the neural network field enjoys a resurgence of interest and a corresponding increase in funding.

A neural network consists of simple processing units and each of the processing units has natural inclination for storing experimental knowledge and making it available for use. These simple processing units, called neurons or perceptions, form distributed network. An artificial neural network is an abstract simulation of a real nervous system that contains a collection of neuron units communication with each other via axon connections. Due to its self-organizing and adaptive nature, the model potentially offers a new parallel processing paradigm that could be more robust and user-friendly than the traditional approaches. As in nature, the network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements.

A neuron is a processing unit, which has n inputs and m outputs. $x_1, x_2, \cdots, x_n$ are outputs of previous layers. w_{ij} is the weight by which neuron i contribute to neuron j. b_j is the threshold of neuron j. The net input net_j is defined by [3]

$$net_j = \sum_{i=1}^{n} x_i w_{ij} - b_j$$

where O_j is the output of the neuron j. Then $O_j = f(net_j)$.

f is a transfer function, which takes the argument input and produces the output. The transfer function is very often a sigmoid function, in part because it is differentiable(Figure 1). The sigmoid transfer function is

$$f(net) = \frac{1}{1+e^{-net}} \tag{5}$$

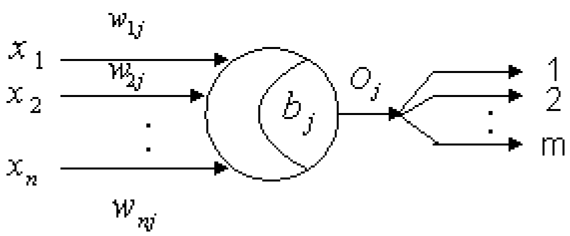

Figure 1. The neuron.

The back-propagation network represents one of the most classical examples of an ANN, being also one of the most simple in terms of the overall design. The network is a straight feedforward network: each neuron receives as input the outputs of all neurons from the previous layer. We adopt a three-layer back-propagation network (see Figure 2). The pretreatment life data are fed to the inputs. The output of network is life distribution. The network has some hidden. The objective is to train the weights and the thresholds, so as to minimize the least-squares-error between the teacher and the actual response.

In this paper, A standard three-layer multi-layer perceptron trained using the back propagation (BP) algorithm is used. The back-propagation network has one input, tree hidden neurons and one output. The value of time is input, and the forecasting value is output. The ANN was trained with the following parameters: learning parameter=0.5, momentum=0.2, error=0.01. The forecasting value data are inputs of trained network. The actual output of network can be calculated by using these weights and the thresholds. We can get the forecasting values of ANN listed in Table 2.

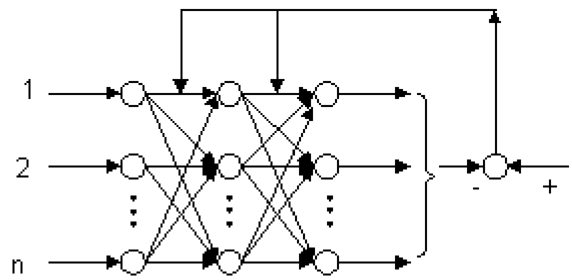

Figure 2. A three-layer Back-propagation network.

TABLE II.
THE ACTUAL TRAFFIC VOLUME FROM 1996 TO 2003 AND THE FORECASTING VALUES OF ANN

Time	Actual Data	forecast value	relative error
1996	984	1012.0	2.85%
1997	977	1026.2	5.03%
1998	976	1040.6	6.61%
1999	990	1055.2	6.59%
2000	1039	1070.2	3.00%
2001	1056	1085.3	2.78%
2002	1116	1100.7	1.37%
2003	1160	1116.4	3.76%
2004		1129.5	
2005		1145.6	
2006		1161.9	
2007		1178.4	
2008		1195.1	

C. Support Vector Machine Predicting Model

Support vector machine(SVM) proposed by Vapnik in 1992 is a new machine 1earning method, which is developed based on Vapnikcher vonenkis (VC) dimension theory and the principle of structural risk minimization(SRM) from statistical learning theory. Originally, SVM were developed for pattern recognition problems. Recently, with the introduction of e-insensitive loss function, SVM have been extended to solve non-linear regression problems. SVM has been tested on a lot of application fields including classification, time, serial estimation, function approximation, text recognition, etc.. SVM has the comprehensive theory foundation such as the universal convergence, speed of convergence, controllability of generalization ability.

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. In simple words, given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. Intuitively, an SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training datapoints of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

SVM models were originally defined for the classification of linearly separable classes of objects. Such an example is presented in Figure 3. For these two-dimensional objects that belong to two classes (class +1 and class-1), it is easy to find a line that separates them perfectly.

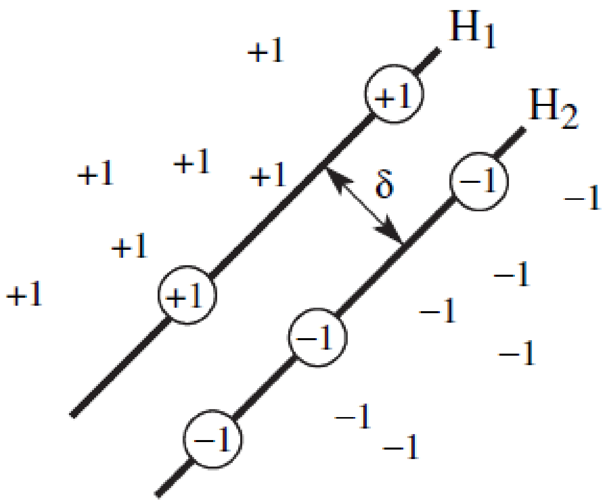

Figure 3. Maximum separation hyperplane.

For any particular set of two-class objects, an SVM finds the unique hyperplane having the maximum margin (denoted with δ in Figure 1). The hyperplane H1 defines the border with class +1 objects, whereas the hyperplane H2 defines the border with class -1 objects. Two objects from class +1 define the hyperplane H1, and three objects from class -1 define the hyperplane H2. These objects, represented inside circles in Figure 1, are called support vectors. A special characteristic of SVM is that the solution to a classification problem is represented by the support vectors that determine the maximum margin hyperplane.

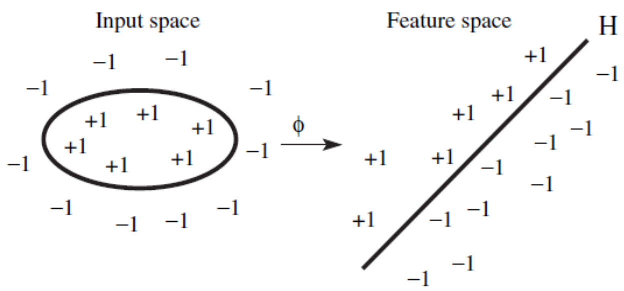

Figure 4. Linear separation in feature space.

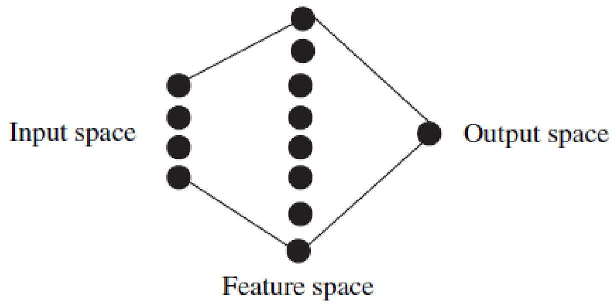

Figure 5. Support vector machines map the input space into a high-dimensional feature space.

SVM[4] can also be used to separate classes that cannot be separated with a linear classifier (Figure 4, left). In such cases, the coordinates of the objects are mapped into a feature space using nonlinear functions called feature functions ϕ. The feature space is a high-dimensional space in which the two classes can be separated with a linear classifier (Figure 4, right).

As presented in Figures 4 and 5, the nonlinear feature function ϕ combines the input space (the original coordinates of the objects) into the feature space, which can even have an infinite dimension. Because the feature space is high dimensional, it is not practical to use directly feature functions ϕ in computing the classification hyperplane. Instead, the nonlinear mapping induced by the feature functions is computed with special nonlinear functions called kernels. Kernels have the advantage of operating in the input space, where the solution of the classification problem is a weighted sum of kernel functions evaluated at the support vectors.

Consider the problem of approximating the set of data, $D = \{(x_i, y_i) \mid i = 1,2,\cdots,l\}$, $x_i \in R^n$, $y_i \in R$, with a linear function[4],

$$f(x) = \langle w, x \rangle + b \tag{6}$$

Using ε -insensitive loss function,

$$L_\varepsilon(y) = \begin{cases} 0 & for \quad |f(x) - y| < \varepsilon \\ |f(x) - y| - \varepsilon & otherwise \end{cases} \tag{7}$$

Support vector machines were extended by Vapnik for regression by using an ε -insensitive loss function (Figure 4). The learning set of patterns is used to obtain a regression model that can be represented as a tube with radius ε fitted to the data. In the ideal case, SVM regression finds a function that maps all input data with a maximum deviation ε from the target (experimental) values. In this case, all training points are located inside the regression tube. However, for datasets affected by errors, it is not possible to fit all the patterns inside the tube and still have a meaningful model. For the general case, SVM regression considers that the error for patterns inside the tube is zero, whereas patterns situated outside the regression tube have an error that increases when the distance to the tube margin increases (Figure 6).

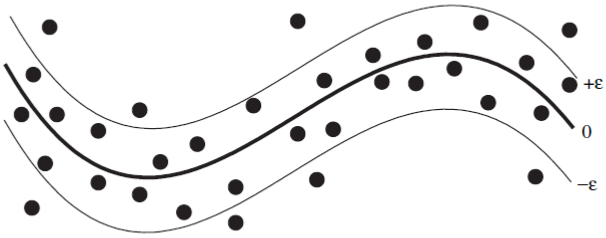

Figure 6. Support vector machines regression determines a tube with radius e fitted to the data.

The SVM regression is depicted in Figure 7. The regression tube is bordered by the hyperplanes $f(x)=\langle w,x\rangle+b+\varepsilon$ and $f(x)=\langle w,x\rangle+b-\varepsilon$. Patterns situated between these hyperplanes have the residual (absolute value for the difference between calculated and experimental y) less than ε, and in SVM regression, the error of these patterns is considered zero; thus, they do not contribute to the penalty. Only patterns situated outside the regression tube have a residual larger than ε and thus a nonzero penalty that, for the ε-insensitive loss function, is proportional to their distance from the SVM regression border (Figure 7, right).

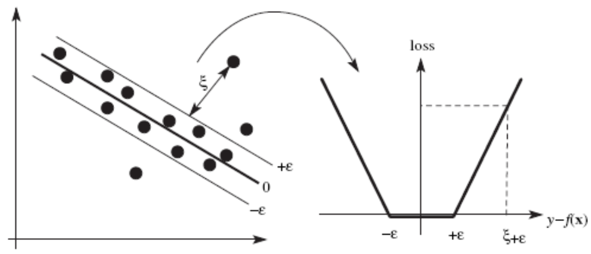

Figure 7. Linear SVM regression case with soft margin and ε-insensitive loss function.

The optimal regression function is given by the minimum of the functional,

$$\min_{w,b,\xi_i,\xi_i^*}\Phi=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)$$
$$s.t.\quad ((w\cdot x_i)+b)-y_i\le\varepsilon+\xi_i\qquad i=1,2,\cdots l \tag{8}$$
$$y_i-((w\cdot x_i)+b)\le\varepsilon+\xi_i^*\qquad i=1,2,\cdots l$$
$$\xi_i,\xi_i^*\ge 0, i=1,2,\cdots l$$

where C is a pre-specified value, and ξ_i, ξ_i^* are slack variables representing upper and lower constraints on the outputs of the system.

Equivalently one can solve the dual formulation of the optimization problem:

$$\max_{\alpha,\alpha^*}W=-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle x_i,x_j\rangle$$
$$+\sum_{i=1}^{l}[\alpha_i(y_i-\varepsilon)-\alpha_i^*(y_i+\varepsilon)] \tag{9}$$
$$s.t.\quad \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0$$
$$0\le\alpha_i,\alpha_i^*\le C, i=1,2,\cdots l$$

Solving Equation (9) determines the Lagrange multipliers α_i, α_i^*, and the regression function is given by

$$w=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)x_i \tag{10}$$

The non-linear mapping can be used to map the data into a high dimensional feature space where linear regression is performed. The kernel approach is again employed to address the curse of dimensionality. The non-linear SVR solution, using ε-insensitive loss function,

$$f(x)=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)K(x_i,x)+b \tag{11}$$

is given by,

$$\max_{\alpha,\alpha^*}W=-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x_j)$$
$$+\sum_{i=1}^{l}[\alpha_i(y_i-\varepsilon)-\alpha_i^*(y_i+\varepsilon)] \tag{12}$$
$$s.t.\quad \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0$$
$$0\le\alpha_i,\alpha_i^*\le C, i=1,2,\cdots l$$

where $K(x_i,x)$ is the kernel function performing the non-linear mapping into feature space. There are many kernel functions, such as polynomial function, radial basis function, exponential radial basis function and multi-layer perception function etc.. Table 3 illustrates the SVR solution for a exponential radial basis function with $C=1000$, $\varepsilon=0.0001$ and $\sigma=18$.

TABLE III.
THE ACTUAL TRAFFIC VOLUME FROM 1996 TO 2003 AND THE FORECASTING VALUES OF SVM

Time	Actual Data	forecast value	relative error
1996	984	984.0	0.00%
1997	977	976.8	0.02%
1998	976	980.4	0.45%
1999	990	994.8	0.48%
2000	1039	1020.0	1.83%
2001	1056	1056.0	0.00%
2002	1116	1102.7	1.19%
2003	1160	1160.0	0.00%
2004		1227.8	
2005		1306.0	
2006		1394.3	
2007		1492.6	
2008		1600. 7	

The exponential radial basis function is given by,

$$K(x_i, x) = \exp(-\frac{\|x_i - x\|}{2\sigma^2}) \qquad (13)$$

We can get the forecasting values of SVM listed in Table 3 and Figure 8. The sum of squares errors of three methods are described in Table4. From Table 4, we can conclude that the accuracy of the support vector machine method is higher than the other two methods.

TABLE IV.
THE FORECASTING VALUES OF THREE METHODS THE SUM OF SQUARES ERRORS OF THREE METHODS

Methods	GM	ANN	SVM
sum of squares errors	1859.8	15587.9	579.7
average relative error	1.18%	4.00%	0.50%
maximum relative error	2.75%	6.61%	1.83%

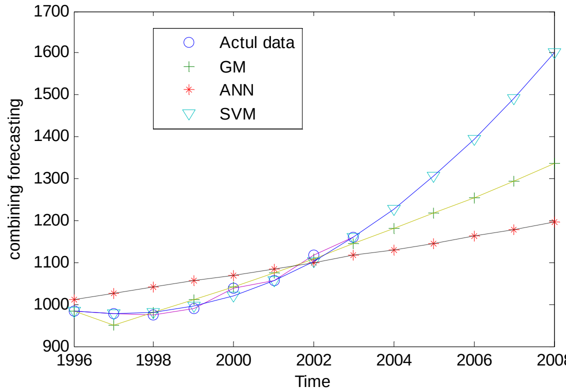

Figure 8. Three forecasting methods

III. COMBINING FORECASTING

A. Linear Combining Forecasting Method

Suppose the real values of some forecasting problem in a period are $X = (x_1, x_2, \cdots, x_n)^T$, and there are m kinds of feasible individual forecasting models around it , whose forecasting values are $X_i = (x_{1i}, x_{2i} \cdots, x_{ni})^T$ ($i = 1,2,\cdots,m$) respectively. Furthermore, let the weighting vector among the m kinds of models be $W = (w_1, w_2, \cdots, w_m)^T$, We denote the values of combining forecasting model are

$$Y = w_1 X_1 + w_2 X_2 + \cdots + w_m X_m$$

The sum of squares error is $Q = (X - Y)^T (X - Y)$. Let $\frac{\partial Q}{\partial w_i} = 0$, $i = 1,2,\cdots,m$, we obtain [5]

$$\begin{cases} w_1 \sum_{i=1}^{n} x_{i1}^2 + w_2 \sum_{i=1}^{n} x_{i1}x_{i2} + \cdots w_m \sum_{i=1}^{n} x_{i1}x_{im} = \sum_{i=1}^{n} x_i x_{i1} \\ w_1 \sum_{i=1}^{n} x_{i2}x_{i1} + w_2 \sum_{i=1}^{n} x_{i2}^2 + \cdots w_m \sum_{i=1}^{n} x_{i2}x_{im} = \sum_{i=1}^{n} x_i x_{i2} \\ \cdots \\ w_1 \sum_{i=1}^{n} x_{im}x_{i1} + w_2 \sum_{i=1}^{n} x_{im}x_{i2} + \cdots w_m \sum_{i=1}^{n} x_{im}^2 = \sum_{i=1}^{n} x_i x_{im} \end{cases} \qquad (14)$$

Combining GM, ANN and SVM, the linear combining forecasting model is $Y = 0.2184 X_{GM} - 0.0591 X_{NN} + 0.8446 X_{SVM}$. The forecasting values of road goods traffic volume is listed in Table 5.

TABLE V.
THE ACTUAL TRAFFIC VOLUME FROM 1996 TO 2003 AND LINEAR COMBINING FORECASTING VALUES(TEN MILLION)

Time	Actual Data	forecast value	relative error
1996	984	986.2	0.22%
1997	977	971.9	0.52%
1998	976	980.6	0.47%
1999	990	998.6	0.87%
2000	1039	1026.0	1.25%
2001	1056	1062.6	0.63%
2002	1116	1108.6	0.67%
2003	1160	1163.7	0.32%
2004		1228.0	
2005		1301.2	
2006		1383.2	
2007		1473.9	
2008		1573.1	

B. Neural network combining forecasting model

The back-propagation network has three inputs (three single method), five hidden neurons and one output., and the forecasting value is output. The ANN was trained with the following parameters: learning parameter=0.5, momentum=0.2, error=0.00001. The forecasting value data are inputs of trained network. The actual output of network can be calculated by using these weights and the thresholds. The forecasting values of the gross of waste water data is listed in Table6.

TABLE VI.
THE ACTUAL TRAFFIC VOLUME FROM 1996 TO 2003 AND NEURAL NETWORK COMBINING FORECASTING VALUES(TEN MILLION)

Time	Actual Data	forecast value	relative error
1996	984	989.91	0.60%
1997	977	977.79	0.08%
1998	976	991.66	1.60%
1999	990	1010.29	2.05%
2000	1039	1033.96	0.49%
2001	1056	1062.89	0.65%
2002	1116	1097.43	1.66%
2003	1160	1137.87	1.91%
2004		1184.12	
2005		1237.59	
2006		1298.2	
2007		1366.45	
2008		1442.85	

C. Support vector machine combining forecasting model

The value of three single methods are inputs in SVM, and the forecasting value is output (Figure 9)[6-8]. Table 7 illustrates the SVR solution for a exponential radial basis function with $C = 1000$, $\varepsilon = 0.0001$ and $\sigma = 960$. The forecasting values of road goods traffic volume is listed in Table 7 and Figure 10. The sum of squares errors of three combining forecasting are described in Table 8. From Table 8, we can conclude that the accuracy of the support vector machine combining forecasting is higher than the other two combining methods.

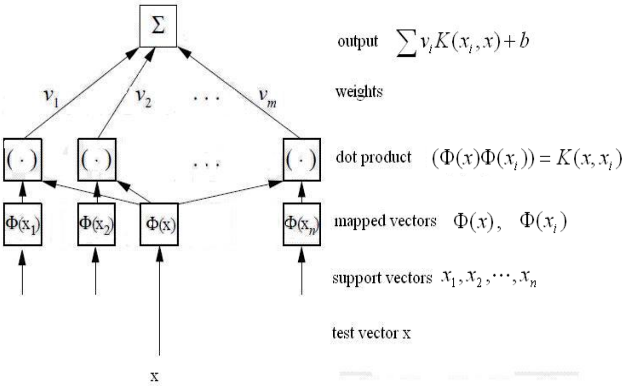

Figure 9. Architecture of a regression machine constructed

TABLE VII.
THE ACTUAL TRAFFIC VOLUME FROM 1996 TO 2003 AND SVM COMBINING FORECASTING VALUES(TEN MILLION)

Time	Actual Data	forecast value	relative error
1996	984	984.0	0.00%
1997	977	976.4	0.06%
1998	976	980.1	0.42%
1999	990	994.6	0.47%
2000	1039	1029.9	0.87%
2001	1056	1056.0	0.00%
2002	1116	1102.7	1.19%
2003	1160	1160.0	0.00%
2004		1227.7	
2005		1305.7	
2006		1393.5	
2007		1491.1	
2008		1598.2	

TABLE VIII.
THE SUM OF SQUARES ERRORS OF TWO COMBINING FORECASTING

Methods	Linear combining forecasting	Neural network combining forecastin	SVM combining forecasting
sum of squares errors	406.98	1599.93	298.03
average relative error	0.62%	1.13%	0.38%
maximum relative error	1.25%	2.05%	1.19%

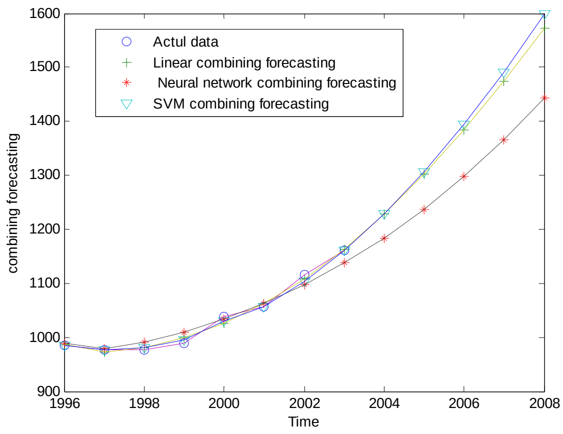

Figure 10. Three combining forecasting

VII. CONCLUSIONS

The use of SVMs in the road goods traffic volume forecasting is studied in this paper. The study has concluded that SVM provide a promising alternative to time series forecasting because they use a risk function consisting of the empirical error and a regularized term which is derived from the structural risk minimization principle. Compared with single prediction methods, linear combining forecasting method and neural network combining forecasting method, the accuracy of the support vector machine method is higher.

ACKNOWLEDGMENT

This work was partially supported by Artificial Intelligence of Key Laboratory of Sichuan Province (2009RY001) and the National Natural Science Foundation of China under Grant No.60773059.

REFERENCES

[1] J M. Bates, C.W.J Granger. Combination of forecasts, Operations Research, pp.451-468,1969， 20(4), pp.451-468.
[2] J. L. Deng. Multidimensional grey planning, Huazhong University of Science and Technology Press, pp.30-35,1990.
[3] L. M. Zhang. Models and applications of artificial neural networks. Shanghai: Fudan University Press, pp.32-47,1994.
[4] S. R. Gunn. Support Vector Machines for Classification and Regression, Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.
[5] X. W. Tang ,C. X. Cao. Study of combination forecasting method, Control and Decision, 1993,8(1). pp.7-12.
[6] W. D. Zhou, L. Zhang, L. C. Jiao. Linear Programming Support Vector Machines. Acta Electronica Sinica, 29(11), pp.1507-1511, 2001.(in Chinese).
[7] J. H. Xu, X. G. Zhang. Nonlinear Kernel Forms of Classical Linear Algorithms. Control and Decision, 1(1), pp.1-6, 12, 2006. (in Chinese).

[8] W. S. An, Y. G. Sun. A New Method for Constructing Kernel Function of Support Vector Regression. Information and Control, 35(3), pp.378-381, 2006. (in Chinese).

Shang Gao was born in 1972, and received his M.S. degree in 1996 and Ph.D degree in 2006. He now works in school of computer science and technology, Jiangsu University of Science and Technology. He is a professor and He is engage mainly in systems engineering and soft computing.

Zaiyue Zhang was born in 1961, and received his M.S. degree in mathematics in 1991 from the Department of Mathematics, Yangzhou Teaching College, and the Ph.D. degree in mathematics in 1995 from the Institute of Software, the Chinese Academy of Sciences. His current research areas are recursion theory, knowledge representation and knowledge reasoning.

Cungen Cao was born in 1964, and received his M.S. degree in 1989 and Ph.D. degree in 1993 both in mathematics from the Institute of Mathematics, the Chinese Academy of Sciences. Now he is a professor of the Institute of Computing Technology, the Chinese Academy of Sciences. His research area is large scale knowledge processing.