Modeling and Forecasting of Urban Logistics Demand Based on Support Vector Machine

Meijuan Gao
Department of Automatic Control
Beijing Union University
Beijing, China
ttgg500@163.com

Qian Feng
School of Economics and Management
Beijing Jiaotong University
Beijing, China
marcie_qfeng@hotmail.com

***Abstract*—Because logistics system was an uncertain, nonlinear, dynamic and complicated system, it was difficult to describe it by traditional methods. The support vector machine (SVM) has the ability of strong nonlinear function approach, it has the ability of strong generalization and it also has the feature of global optimization. In this paper, a modeling and forecasting method of urban logistics demand based on regression SVM is presented. The SVM network structure for forecasting of urban logistics is established. Moreover, we propose a self-adaptive parameter adjust iterative algorithm to confirm SVM parameters, thereby enhancing the convergence rate and the forecasting accuracy. With the ability of strong self-learning and well generalization of SVM, the modeling and forecasting method can truly forecast the logistics demand by learning the index information of affect logistics demand. The actual forecasting results show that this method is feasible and effective.**

Keywords- urban logistics; demand; modeling and forecasting; support vector machine

I. INTRODUCTION

As an important part of the national economy and the most economic mode of service in the process of industrialization, the logistics industry is growing rapidly on a global scale. In recent years, along with the rapid development of city economic, it has an important strategic meaning to develop the modern logistics for optimizing the economic structure, improving the investment environment and enhancing the competitiveness of the urban economy as a whole. Therefore, to study the logistics planning has become an important issue. The process of logistics planning is the planning and decision-making process in fact, and the successful planning and decision-making depends on the accuracy of the forecast. Therefore, the forecast is the key link related to the success of logistics planning. That is, the primary issue in planning the logistics is the logistics demand forecast [1].

Logistics forecast that is "Investigating and researching on the logistics flow direction, flow rate, cash turnover and the law of supply and demand etc to access to the variety of data and information, then expecting and forecasting the state of logistics in a certain period with scientific methods." Nowadays, there are many ways to forecast the logistics, such as moving average method, exponential smoothing method, time series decomposition method, regression model method etc. These traditional methods are required to establish a clear functional relation with the forecasted goal, but the logistics system is always an open complex systems. There are a number of factors including both qualitative factors and quantitative indicators impact of the changes in the demand. In addition, the development of the logistics demand is always presented in a general non-linear and random state, and it is closely related to the level of economic development of the city, the government policies and other factors etc. These factors are not only difficult to describe quantitatively, but also difficult to establish the certain functional relation with the forecasted goal [2]-[3]. So, it is not appropriate to use the aforementioned methods.

The BP network based on gradient descend is a new method in recent years, its ability to approach nonlinear function has been proved in theory also have been validated in actual applications [4]. But the BP network has some problems such as converge to local minimum, the overfitting, convergence rate slowly and the network structure is always decided by experience because it doesn't have a good guiding theory.

Support Vector Machine (SVM) proposed by Vapnik is a newly developed technique which based on statistical learning theory [5]-[6], it adopts Structure Risk Minimization principle which avoids local minimum and effective solves the over learning and assures good generalization ability and better predict accuracy. The special predominance of SVM in resolve limited samples, non-linear function and multidimensional pattern recognition make it become a kind of excellent machine learning method [7]. Considering the issues that the logistics system is a complicated and nonlinear system, the logistics system with a small sample of data and a lack of internal relations and laws, and the special predominance of SVM possess, the regression SVM is applied to the forecasting of urban logistics demand in this paper.

II. METHOD STUDY

A. *The Principle of SVM*

SVM is proposed which is based on the idea of optimal classify hyperplane of linearly separable. Suppose we have two classes samples of linearly separable, H is the class line which divide the two classes without mistake, H_1 and H_2

978-0-7695-3543-2/09 $25.00 © 2009 IEEE
DOI 10.1109/WKDD.2009.211

are the line that pass through the points which are the nearest to the class line in each classes samples and parallel to the class line. The distance between H_1 and H_2 is called the separating margin of the two classes. We want the optimal class line not only can separate the two classes correctly which ensure the experience risk minimization, but also can have the maximum separating margin of the two classes which ensure the real risk minimization. For the high dimension, the optimal class line is the optimal classify hyperplane.

The classify function can be described as follow:

$$f(x) = \operatorname{sgn}\{(w^* \cdot x) + b^*\} = \operatorname{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^*\} \quad (1)$$

Here, w is the normal of the hyperplane, b decides the place which relative the origin point, (x_i, y_i) is linearly separable sample set.

For the optimal non-linear classify problem, we can transform it to another high dimensional space such that the problem will be linear problem. If the dot product $(x \cdot x_i)$ is replaced by the kernel function $K(x, x')$, equal that transforms the original characteristic space to a new characteristic space, the optimal classify function as follow:

$$f(x) = \operatorname{sgn}(\sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x) + b^*) \quad (2)$$

The rest condition of the algorithmic keep changelessness, this is Support Vector Machine.

The classify function obtained by the support vector machine is analogous to a neural network in formally, the sketch map of support vector machine is shown in Fig. 1.

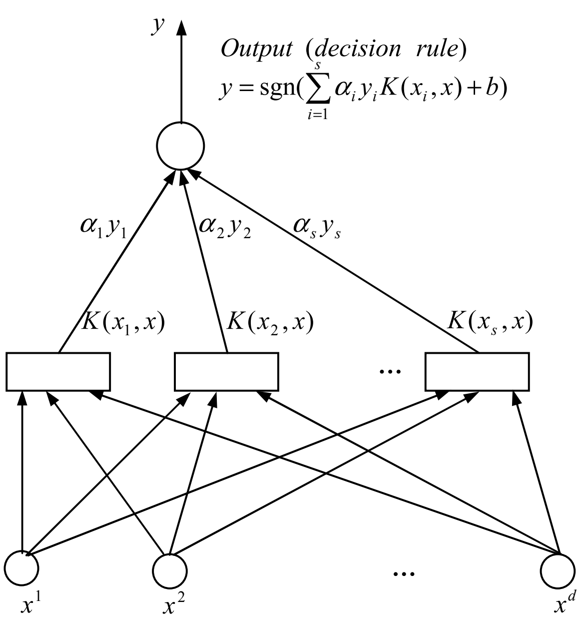

Figure 1. The sketch map of support vector machine

The basic thought of the SVM could be generalized as following: at first, transform the input space to a high dimensional space by non-linear transformation, then we work out the optimal classify hyperplane in the new space, while the non-linear transformation is achieved by defining a proper kernel function.

B. *Regression Support Vector Machine*

At first, the mathematics description of the function regression problem is present. Suppose have the sample data $\{x_i, y_i\}$ ($i = 1, \cdots, l$). Among them, $x_i \in R^n$ are n-dimensional sample input, $y_i \in R$ are sample output, then the function regression is to find a function f through sample training, except for the training sample, x can find out the corresponding y through f.

Definition 1 (linearity) ε -insensitive loss function $L^{\varepsilon}(x, y, f)$ is defined as follow:

$$L^{\varepsilon}(x, y, f) = | y - f(x) |_{\varepsilon} = \begin{cases} 0 & | y - f(x) | \le \varepsilon \\ | y - f(x) | - \varepsilon & else \end{cases} \quad (3)$$

Here, f is a real value function on the field X. The new loss function describes a sort of ε insensitivity model that if the difference between predicted value and actual value is less than ε, the loss equal to 0.

Through solve the optimization problem (4) and (5), we can get w and the estimate function (6).

$$\min \phi(w) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \hat{\xi}_i) \quad (4)$$

$$subject\ \ to \begin{cases} y_i - ((w \cdot x_i) + b) \le \varepsilon + \xi_i \\ ((w \cdot x_i) + b) - y_i \le \varepsilon + \hat{\xi}_i \\ \xi_i, \hat{\xi}_i \ge 0 \quad i = 1, \cdots, l \end{cases} \quad (5)$$

Here, $C (C > 0)$ is a constant called penalty factor which be used to express the compromise between the smoothness of function f and the value of that allow error is greater than ε. $\xi, \hat{\xi}$ are slack variable.

$$\begin{cases} w = \sum_{i=1}^{l} (\alpha_i - \hat{\alpha}_i)\ x_i \\ f(x) = \sum_{i=1}^{l} (\alpha_i - \hat{\alpha}_i)\ (x_i, x) + b \end{cases} \quad (6)$$

Here, x_i which correspond to $(\alpha_i - \hat{\alpha}_i) \ne 0$ is support vector.

We use kernel function $k(x, x')$ to replace dot product in the formula (6), and obtain the regression SVM function (7).

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \hat{\alpha}_i) k(x_i, x) + b \quad (7)$$

Here, kernel function $k(x_i, x)$ is discretional symmetry function which satisfy the Mercer condition.

III. URBAN LOGISTICS DEMAND AFFECT FACTORS

At present, the indicator system of measuring logistics demand includes the index of physical quantity and magnitude of value. The index of physical quantity mainly includes freight volume, freight turnover, inventory, processing capacity etc. The index of magnitude of value mainly includes logistics cost, logistics income, supplied chains' increment etc. Due to lacking of necessary statistical data, the method of magnitude of value commonly processes empirical forecasting for logistics demand based on expert experience. Hence, according to attainability of datum, the physical quantity system is used to do a quantitative research on logistics demand.

In all kinds of logistics activities, transportation is a most fundamental factor. The main indicators of transportation embody on the data of freight volume and freight turnover etc. The freight volume is used to represent logistics demand in this paper. Although the freight volume can't represent all the logistics quantity of work, but transport is the most basic activities on logistics process, it runs through whole logistics process. Therefore, using freight volume to represent logistics demand can reflect its change law to some extent.

The purpose of logistics forecasting is estimating the future development of the logistics. There are many factors influencing logistics forecasting, such as market supply and demand situation, economic and traffic etc, at the same time, these factors are the content of logistics forecasting, and various factors are mutual dependence and mutual restraint. Therefore it is a complex system engineering to establish the logistics forecasting model. We should choose logistics index based on the principle of comprehensiveness, representativeness scientificalness and acquisitiveness. So there are two index systems which can reflect urban logistics demand according to the principle.

The first kind index is the economy index relate to logistics. Because logistics demand lies on urban scales, urban total output value of industry and agriculture, urban total quantity consumed. So the factors reflecting freight volume mainly include the number of total population, the resident consumption level, primary industry, secondary industry, tertiary industry, the total output value of production, the total output value of industry and agriculture, the total retail amount of commodities.

The second kind is the other factors relate to logistics, such as macroeconomic policy, the change of market environment, the change of consumption concept, technical progress and logistics service levels.

IV. URBAN LOGISTICS DEMAND MODELING BASED ON REGRESSION SVM

A. Parameters Ascertain of Regression SVM

The performance of support vector machine method depends on the selection of kernel function and its width coefficient, penalty factor C and insensitive coefficient ε, so the preferences is very important, it will influence the result of regression directly.

We adopt regression SVM of radial basis function (RBF). The insensitive coefficient ε reflects the limitation of SVM allowed noise range in data. The insensitive coefficient ε adopts $\varepsilon = 0.5D$ in this paper, and D is the variance of noise distributing function. We proposed a self-adaptive parameter adjust iterative algorithm to confirm penalty factor C and kernel function width parameter σ^2. The steps of iterative algorithm as follow:

(1) Confirm the initial value of penalty factor C and kernel function width parameter σ^2 randomly, execute support vector regression estimate using the initial values, and compute the average fitting relative error.

(2) Adjust penalty factor C and width parameter σ^2 according to certain step length ΔC and $\Delta\sigma^2$, execute support vector regression estimate using the adjusted parameter, and compute the average fitting relative error of the adjusted parameter.

(3) If the average fitting relative error is smaller and it can satisfy the requirement of system, them we should judge the complexity of the regression function, if it was not complex, turn(4); if the average fitting relative error is not smaller, then adjust step length ΔC and $\Delta\sigma^2$, turn (2); if the average fitting relative error is smaller, but it can not satisfy the requirement of system, turn (2); if average fitting relative error can satisfy the requirement of system, but the regression function is too complex, then adjust step length ΔC and $\Delta\sigma^2$, and turn (2).

(4) Get the optimal regression SVM parameters and the iterative algorithm end.

B. Application Study

Because the second kind index relate to urban logistics demand such as macroeconomic policy, the change of market environment etc can't be quantified, so they shall not be considered in this paper. We only consider the first kind economy index relate to urban logistics demand and use freight volume to represent the urban logistics demand.

Through the correlative analysis between the affect factors of urban logistics demand and urban freight volume, we find the affect factors which have the maximum correlativity with the urban freight volume. Combining expert experience, finally, we select 7 kinds of characteristic parameters to as input vectors of regression SVM, these parameters are urban GDP, output value of primary industry, output value of secondary industry, output value of tertiary industry, urban total volume of retail sales, total export import volume and per capita consumption level. The output vector of regression SVM represents the urban freight volume. In mathematical forecasting model, Y_i is the urban logistic demand vector (freight volume) of the year i, $X_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ is the economic factors that effect urban logistics demand, the x_{in} is the variable value of the n economic factor in i year. Because of the economic effect factors of urban logistics demand will not be drastic changes by leaps and bounds in the short term, so the economic effect factors of the $i-1$ year can be used to forecast the urban logistics demand of the i year. In this way, put $(x_{(i-1),1}, x_{(i-1),2}, \cdots, x_{(i-1),n})$ as input variables of regression SVM, and the Y_i is used as the output variables of regression SVM, We can get many group training samples and testing samples and every training sample is composed of 7-input and 1-output. We can obtain the dynamic change relation between the urban logistics demand and economic effect factors by using the training samples to train the regression SVM network and can obtain the forecasting value of 2008 year urban logistics demand by input the data of 2007 year according the regression SVM network model.

We write the algorithm program using C++. In order to show the advantage and feasibility of regression SVM method, we adopt regression SVM and radial basis function neural network (RBFNN) to forecast the urban logistics demand (freight volume), SVM uses radial basis function, the best parameters of SVM through the self-adaptive parameter adjust iterative algorithm optimize is $\varepsilon = 0.1$, C =200, σ^2=2, for the RBFNN, the system error is 0.001, the width of kernel function σ^2 is 2, equal to the width of kernel function of SVM. After all samples are normalized, we select different training sample numbers and different test samples numbers every time and input the training samples into the regression SVM model and RBFNN to learn, then use the test samples to test the result. The result of train and test by regression SVM and RBFNN is shown in table 1.

TABLE I. TRAIN AND TEST RESULTS BY SVM AND RBFNN

train sample	test sample	Regression SVM		RBFNN	
		mean squared errors of train sample	mean squared errors of test sample	mean squared errors of train sample	mean squared errors of test sample
20	4	0.0098	0.0081	0.0068	0.1037
30	6	0.0085	0.0093	0.0056	0.1729
40	8	0.0069	0.1036	0.0045	0.3982
50	10	0.0052	0.1259	0.0032	0.5216

From the table 1, we can see the mean squared errors of training samples of SVM is bigger than the RBNN, but to the mean squared errors of testing samples, the RBFNN is bigger. It shows that the generalized ability of the regression SVM is stronger than the RBFNN. When the numbers of the training sample were changed, the fluctuating range of the generalize error of SVM is smaller than the RBFNN, it shows that the degree of depending on the sample data of regression SVM is smaller than the RBFNN. So the model of urban logistics demand based on regression SVM has higher stability, and can obtain higher forecasting accuracy.

The reason of generation above result is that the traditional method such as RBFNN is based on Experience Risk Minimization (ERM) principle and uses the error back propagation method, but the ERM is difficult to prove the Real Risk Minimization, this is the reason that why the ANN easy appears overfitting. The support vector machine is based on Structure Risk Minimization principle, the optimization of objective function includes two indices that are Experience Risk and confidence interval, and both of them decide the Real Risk of SVM together. Therefore the predictive ability of SVM is stronger than the traditional neural network, in addition, the small sample learning peculiarity of SVM also decide that the degree of depending on the sample data of SVM is smaller than the RBFNN.

V. CONCLUSION

The forecast of urban logistics demand is to find the internal relationship between urban economy and urban logistics demand and to provide necessary decision-making data for urban logistics planning. In this paper, the regression SVM is used to establish the urban logistics demand model, it not only reveals the internal non-linear mapping relationship between urban economy and urban logistics demand, but also puts forward a new idea and methodology for urban logistics demand forecasting. Owing to the SVM is a convex quadratic optimization problem, it can assure the extremum result is the global optimum result, and it can effectively solve the overfitting problem of artificial neural network [8], Therefore, the modeling method based on the regression SVM can enhance the converging speed and the forecasting accuracy of urban logistics demand (freight volume) to a great extent. The real forecast result shows that the method can truly forecast the urban logistics demand, and it can reflect the objective development trend of logistics demand with more reliable accuracy.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of Chin under Grant No.40674028 and the Beijing Education Special Project under Grant PXM2008_014209_054622.

REFERENCES

[1] Xiao Dan, Ni Mei, Li Yisong, "Researching on the Analytic Index of Logistics Demand", Modern logistics, no. 21, pp. 33-24, 2003.

[2] Cai Xiaoli, Chen Chouyong, "The Prediction and Analyse of Regional Logistics Demand", Logistics Sci-tech, no. 27, pp.15-18, 2004.

[3] Xu Youli, "Forecasting LogisticsDemand Based on ArtificialNeuralNetwor", Journal of Zhejian Shuren University, vol. 8, no. 1, pp. 56-58, 2008.

[4] Jiao Licheng, Neural network system theory, Xian: Xi an electronic science and technology University press, 1995.

[5] V.Vapnik, Statistical Learning Theory, Wiley, 1998.

[6] J.Weston and C. Watkins, Support vector machines for multi-class pattern recognition, In Proceeding of the 6th European symposium on Artificial Neural Network (ESANN), 1999.

[7] Nello Cristianini, John Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Beijing: electronic industry publishing house, 2005.

[8] Bian Zhaoqi, Zhang Xuegong, Pattern recognition, Beijing: Tsinghua university publishing house, 2000.

[9] Tang Weihong, Li Wenfen, "Application of Support Vector Machines Based on Time Sequence in Logistics Forecasting", Logistics Sci-tech, vol. 28, no. 113, pp. 8-11, 2005.