Private Car'S Ownership Forecasting Based on Support Vector Regression

Liping Wang, Junchao Yang

School of Economic and Management Administration,
North China Electric Power University, Beijing 102206, China

Abstract
Private car is a main pollution source for the air pollution and the haze phenomenon in big city. How to control the private car's ownership is an important way to reduce the vehicle exhaust emissions in the future, and the forecasting of the private car's ownership plays an important role in make a decision of controlling private car's industry plan, investment and operation. Support vector regression forecasting model is an effective forecasting model of small sample; therefore, this paper uses the support vector regression model to forecasting the private car's ownership, and an experiment results show that the proposed model gets the maximum mean absolute percentage error (MAPE) value 1.284%, which is better than the results of regression, BPNN and SVR.
Key words: PRIVATE CAR'S OWNERSHIP; FORECASTING; SUPPORT VECTOR REGRESSION

1. Introduction

Recently, Haze phenomenon, especially in Beijing, as a serious environment problem has attracted the attention of China. According to the published record from Chinese Academy of Sciences, the vehicle's exhaust is one of the major sources of the air pollution in Beijing. With the economic development of China, the private car, which can bring to people's convenience living conditions, will remains to increase faster year by year, and the private car will pay much attention to the rich households, especially in a big city such as Beijing and Shanghai in China. Plenty of the COx, SOx and other pollution particles with the car air pollutions are emitted by burn the petroleum. Private cars has been become a serious problem which should be immediately controlled. Controlling vehicle exhaust emissions for controlling urban air pollution has become an important task of a city's sustainable development. According to Ministry of Public Security Traffic Management Bureau Statistics, at the end of Jun 2012, the national vehicle has reached 114 million, and there are 17 cities has been reached over one million cars in an annual, in which Beijing, Chengdu, Tianjin, Shenzhen, Shanghai five cities of vehicle ownership are more than two million. Vehicle industry and its ancillary industries have brought great convenience and opportunity to China's economy, but at the same time, the big consumes fossil energy used by vehicle also caused serious air pollution problem. If China wants to control vehicle pollution of the air pollution, it is essential to reform the vehicle industry. The primary problem is slowing down the fuel vehicle demand increasing trend; it is a shortcut of using electric vehicle to instead the fuel vehicle.

Formulating the electric vehicle development plan needs to forecast private cars' ownership precisely. It is important for electric car industry's planning, demand analysis and policy reference. Some scholars have been studied the ownership's forecasting problem. They uses artificial neural network[1-2], Particle Swarm Optimization Algorithm[3] and Time Series forecasting[4] method to forecast car ownership or electric vehicle's number. Other researches study the relationship between the social or environment influence and the hybrid or electric cars[5-7]. However, it lacks using support vector regression forecasting method, which was invented by Vapnik and has better performance in annual forecasting, to study the ownership of private cars.

SVR model is suit for small sample forecasting problem such as the situation of the private cars' ownership forecasting, the SVR model's performance is also better than the others traditional and intelligence forecasting model. In this paper, we will use support vector regression to study this problem, and as the result, It can be seen that SVR forecasting model has the most effectiveness in the private cars' ownership forecasting.

2. Support vector regression Method

Support vector Machine (SVM), based on machine statistical learning theory, which is firstly used to pattern analysis. It has been successfully applied in many areas such as text categorization, pattern recognition, signal processing and so on[8]. With the development of SVM, it has promoted to deal with the non-linear regression function fitting problem, which changed into support vector regression model (SVR) method. Suppose a set of data $G=\{(x_i,d_i)\}, i=1\cdots N$, $x_i \in R^n$ is a n dimension input vector, $d_i \in R^1$ is a corresponding target output, N expresses the total number of pattern records. The linear regression estimate function can be express as[9]

$$y = f(x) = w\psi(x) + b \tag{1}$$

In which, $\psi(x)$ is a nonlinear mapping from the input space to a high dimensional feature space. w is a weight vector and b is a threshold value, which can be estimated by minimizing the regularized risk function.

$$R(C) = (C/N)\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \|w\|^2/2 \tag{2}$$

In which,

$$L_\varepsilon(d,y) = \begin{cases} 0 & |d-y| \le \varepsilon \\ |d-y| - \varepsilon & \text{otherwise} \end{cases} \tag{3}$$

$\|w\|^2/2$ measures the flatness of the function, and the function $L_\varepsilon(d,y)$ is called the ε-insensitive loss function. C is considered to specify the trade-off between the empirical risk and the model flatness, ε express the Vapnik's linear loss function zone to measure empirical error. After two slack variables ζ and ζ^* are introduced to represent the distance from actual values to the corresponding boundary values of ε-tube, Equation (2) can be transformed to Equation (4).

$$\begin{aligned} & R(w,\zeta,\zeta^*) = \|w\|^2/2 + C\sum_{i=1}^{N}(\zeta_i + \zeta_i^*) \\ & s.t. \\ & w\psi(x_i) + b_i - d_i \le \varepsilon + \zeta_i^*, i = 1,2,\ldots,N \\ & d_i - w\psi(x_i) - b_i \le \varepsilon + \zeta_i, i = 1,2,\ldots,N \\ & \zeta_i, \zeta_i^* \ge 0, i = 1,2,\ldots,N \end{aligned} \tag{4}$$

This constrained optimization problem can be solved with the primal Lagrangian form

$$\begin{aligned} & L(w,b,\zeta,\zeta^*,\alpha_i,\alpha_i^*,\beta_i,\beta_i^*) \\ & = \|w\|^2/2 + C\sum_{i=1}^{N}(\zeta_i+\zeta_i^*) - \sum_{i=1}^{N}\beta_i[(w\psi(x_i)+b-d_i+\varepsilon+\zeta_i)] \\ & -\sum_{i=1}^{N}\beta_i^*[(d_i - w\psi(x_i) - b + \varepsilon + \zeta_i^*)] - \sum_{i=1}^{N}(\alpha_i\zeta_i + \alpha_i^*\zeta_i^*) \end{aligned}$$

(5)

w, b, ζ, ζ^* need be minimizedin Eq.(5), therefore, it needs satisfy the following conditions.

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \to w - \sum_{i=1}^{N}(\beta_i - \beta_i^*)\psi(x_i) = 0 \\ \dfrac{\partial L}{\partial b} = 0 \to \sum_{i=1}^{N}(\beta_i - \beta_i^*) = 0 \\ \dfrac{\partial L}{\partial \zeta} = 0 \to C - \zeta - \alpha_i = 0 \\ \dfrac{\partial L}{\partial \zeta^*} = 0 \to C - \zeta^* - \alpha_i^* = 0 \end{cases} \tag{6}$$

Using Karush-Kuhn-Tucker conditions and substituting Eq.(6) in Eq.(5), the dual can be obtained as

$$\vartheta(\beta_i,\beta_i^*)=\sum_{i=1}^{N}d_i(\beta_i-\beta_i^*)-\varepsilon\sum_{i=1}^{N}(\beta_i-\beta_i^*)$$
$$-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\beta_i-\beta_i^*)(\beta_j-\beta_j^*)K(x_i,x_j) \quad (7)$$
$$s.t:\sum_{i=1}^{N}(\beta_i-\beta_i^*)=0,$$
$$0\le\beta_i\le C, 0\le\beta_i^*\le C, i=1,2,\ldots,N.$$

In which, $\beta_i\beta_i^*=0$ and an optimal desired weight vector of the regression hyperplane is

$$w^*=\sum_{i=1}^{N}(\beta_i-\beta_i^*)\psi(x). \quad (8)$$

The final regression function is

$$f(x,\beta,\beta^*)=\sum_{i=1}^{l}(\beta_i-\beta_i^*)K(x,x_i)+b. \quad (9)$$

Where $K(x_i,x_j)$ is the inner product of two vectors in the feature space $\psi(x_i)$ and $\psi(x_j)$, $K(x_i,x_j)=\psi(x_i)\psi(x_j)$ is called the kernel function which needs to meet Mercer's condition. The most commonly used kernel function is the Gaussian RBF kernel function $K(x_i,x_j)=\exp(-\|x_i-x_j\|^2/2\sigma^2)$ (σ is a constant),it is also employed in this study.

3. An empirical example

According to the statistic data about China's car industry Yearbook, the ownership number of private car can be obtained from 1999 to 2012, which is shown in Table 1.As one of the traditional forecasting method, the regression model is chosen as one of a constitution of the comparing forecasting model. In addition, a model of grey forecasting model GM(1,1) is pitched on in consideration of the others comparing model. In SVR forecasting model, the input variables are also expressed as x_{t-3}, x_{t-2} and x_{t-1}, and the output node is also x_t, and its parameters use the default values of libsvm software package in matlab. The forecasting results are shown in Table 1.

Table 1. The actual value and forecasting results of the private car's ownership in China

Year	The forecasting results			
	Actual value	Regression	GM(1,1)	SVR
2000	365.09	--	--	--
2001	469.85	558.134	631.6766	--
2002	623.76	685.194	797.1635	--
2003	845.87	871.867	1006.005	843.6078
2004	1,069.69	1141.258	1269.558	1067.519
2005	1,383.93	1412.722	1602.157	1386.165
2006	1,823.57	1793.855	2021.891	1829.026
2007	2,316.91	2327.081	2551.587	2319.145
2008	2,880.50	2925.439	3220.052	2885.269
2009	3,808.33	3609.001	4063.643	3810.726
2010	4,989.50	4734.339	5128.238	4987.238
2011	6,237.46	6166.945	6471.735	6234.573
2012	7,637.87	7680.559	8167.202	7634.73
2013	9,198.23	9379.076	10306.85	9196.059
MAPE	--	2.680%	10.761%	1.284%

4. Conclusions

SVR model is suit for small sample forecasting problem, in the private cars' ownership forecasting, the SVR model's performance is better than the others, such as regression forecasting model and GM(1,1) forecasting model. It can be concluding that SVR forecasting model has the most effectiveness in the private cars' ownership forecasting.

References

1. Azadeh A, Neshat N, Rafiee K, Zohrevand A.M.(2012) An adaptive neural network-fuzzy linear regression approach for improved car ownership estimation and forecasting in complex and uncertain environments: the case of Iran. *Transportation Planning and Technology*, 35(2), p.p. 221-240.

2. Wu, J.D., J.C. Liu. (2012) A forecasting system for car fuel consumption using a radial basis function neural network. *Expert Systems with Applications*, 39(2), p.p. 1883-1888.
3. Zheng X., Wang M. M., Ma W. M., Xin C. L. (2012) Research on Forecasting the Car Sales in Mainland of China Based on Improved Particle Swarm Optimization Algorithm and Gray Model. *Information-an International Interdisciplinary Journal*, 2012. 15(4), p.p. 1461-1476.
4. Pandey, P., S. Kumar, S. Shrivastav. (2013) Forecasting using Fuzzy Time Series for Diffusion of Innovation: Case of Tata Nano Car in India. *National Academy Science Letters-India*, 36(3), p.p. 299-309.
5. Orbach Y, Fruchter G.E.(2011) Forecasting sales and product evolution: The case of the hybrid/electric car. *Technological Forecasting and Social Change*, 78(7), p.p. 1210-1226.
6. Pierdzioch, C., Rulke J.C., Stadtmann G..(2011) Forecasting US car sales and car registrations in Japan: Rationality, accuracy and herding. *Japan and the World Economy*, 23(4), p.p. 253-258.
7. Lee, D.H., Park S.Y, Hong J.C, Choi S.J., Kim J.W.(2013) Analysis of the energy and environmental effects of green car deployment by an integrating energy system model with a forecasting model. *Applied Energy*, Vol. 103, p.p. 306-316.
8. Wang, Y.Q., Liu, X.(2015) Improved support vector clustering algorithm for color image segmentation. *Engineering Review*, 35(2), p.p. 121-129.
9. Wang J.J, Li Li, Niu D.X, Tan Z.F.(2012) An annual load forecasting model based on support vector regression with differential evolution algorithm. *Applied Energy*, Vol. 94, p.p. 65-70.