An optimal combined forecasting method to prediction of ownership for private cars

Junchao Yang*

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

Received 1 June 2014, www.cmnt.lv

Abstract

With the high speed economic development of China, the number of private cars is also increasing rapidly. However, car emission has an important influence for air pollution; it needs to consider controlling private cars demand when the government makes automotive industry's strategy. It should forecast the future private car accuracy. In this article, a combined forecasting model with simulate anneal algorithm optimizing weights of private car is carried out. The proposed model can improve the performance of each single forecasting model such as regression, Grey and SVM. Finally, a case study with Chinese private cars number is presented, and the results are shown that the proposed model is superior to each single model.

Keywords: Ownership Demand Prediction, Combined Forecasting, Simulation Anneal Algorithm

1 Introduction

In recent years, with the growing improvement of people's living conditions, car as a convenience transport tool is also increasing rapidly. Up to now, a households, living in a big city, especially in Beijing, Shanghai, Guangzhou and so on, almost has a car. The cars industry helps people shorten their travel time, improve the traffic efficiency and boost national economic development, however, the consumer fossil energy by cars are also producing the air pollution gas, which is attracting the attention of many people. Recently, fog and haze weather frequent appears in Beijing after 2013, it gives the health of people a great threat, one of the reasons is mainly by the car's gas emissions, which contain solid particles, COx, SOx and other pollution particles. The fuel cars are changing an air pollution source threat of a city's sustainable development.

Personal private cars need to control to reduce the air pollution, and the number of ownership is a key problem to deal with. It is important for 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]. Other researchers build a model to predict the cars market size. However, all of these studies try their best to use a single forecasting model to deal with the car's ownership or market size problem. Consequently, it is very important to choose a combined model to enhance the forecasting accuracy. The combined forecasting model option is generally composed of two or more forecasting models. By means of endowing each model with certain weight, the combined forecasting model is regarded as an ideal option to solve the accuracy problem as described above.

As the theoretically prove, a combined model's accuracy should higher than each single model which is included in the combined model. This paper propose a combined model with simulate anneal algorithm optimizing weights to forecast the private car ownership number of China. And the combined model contains the traditional forecasting model like regression, Grey and the intelligence forecasting model like SVM.

2 A combined forecasting model with simulate anneal algorithm

2.1 DESCRIPTION OF COMBINED FORECASTING MODEL

Suppose that f_i represents predictive results of the i method and w_i stands for the weight of i method. Then, the combined forecasting model can be expressed by:

$$\hat{y} = \sum_{i=1}^{n} w_i f_i \tag{1}$$

where, y stands for the actual data and an optimization problem of the combined model needs to satisfy the following Equation 2.

$$Min: MSE(y - \sum_{i=1}^{n} w_i f_i) \tag{2}$$

$$s.t. \sum_{i=1}^{n} w_i = 1; w_i \geq 0$$

The key factor of the combined model is to determine the weight. The methods of setting up weight coefficient, according to the rule whether the weight is the function of time or not, can be divided into fixed weight method and transformable weight method. Owing to the stationary of weight coefficient, the fixed weight determined model is identified as the most universal and popular method in reality. In addition, this method is suitable for using intel-

* ***Corresponding author's*** e-mail: gs_windboy@126.com

ligence algorithm such as genetic algorithm, particle swarm optimization to solve the problem of the weight assignment.

2.2 OPTIMIZING WEIGHTS WITH SIMULATE ANNEAL ALGORITHM

Simulated annealing (SA) algorithm is an iterative solution strategy on the random search algorithm; it is based on the physical annealing process of solid material and the general similarity of combinatorial optimization problems. The name and inspiration come from annealing in metallurgy, a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects. The heat causes the atoms to become unstuck from their initial positions and wander randomly through states of higher energy; the slow cooling gives them more chances of finding configurations with lower internal energy than the initial one. The principle can be described as follows:

At *t* temperature, a new state *j* generated by the current state *i*, and then energy is E_i and E_j, respectively, if $E_i < E_j$, the current state *j* is accepted, otherwise, it need consider whether the state is the "important state", it determined by the probability. If the probability $p_r = \exp(-(E_i - E_j)/kt) > random[0,1]$, (*k* is the Boltzmann constant, random [0,1] expresses that the [0,1] is generated by a probability distribution of random numbers), it is still accepting new state for the current state *j*, otherwise, accepting state *i*. Repeat this process, the system will tend to stay at the lowest energy at the temperature.

The basic simulated annealing (SA) algorithm is as follows:

Step 1. Initialization. Given the scope of model for each parameters, randomly selected an initial solution x_0, and calculate the corresponding target value $E(x_0)$; set the initial temperature T_0, final temperature T_f, make a random number $\varepsilon \in (0,1)$ as a probability threshold, set the cooling function $T(t+1) = \gamma \bullet T(t)$, in which, γ is annealing coefficient, t is the number of iterations.

Step 2. At a certain T temperature, make a perturbation Δx, then a new solution is $x' = x + \Delta x$ produced, calculate the difference $\Delta E(x) = E(x') - E(x)$.

Step 3. If $\Delta E(x) < 0$, x' is accepted; if $\Delta E(x) > 0$, x' is accepted by probability $p = \exp(-\Delta E / K' \bullet T)$, K' is a constant and usually taken the value 1. If $p > \varepsilon$, x' is accepted. When x' accepted, $x' = x$.

Step 4. In a certain temperature, repeat steps 3.

Step 5. Reduce the temperature T by slow cooling function.

Step 6. Repeat steps 2 to step 5, until the condition is meet.

Base on the principle of the simulation anneal algorithm, a novel method of Combining simulate anneal method with optimal a combined forecasting weights in car's ownership prediction is proposed as follows:

Step 1. Initialization

Using simulated annealing algorithm to randomly give an initial weight of the combined model, set the initial temperature $T_0 (T_0 > 0)$, the basic steps V_0, testing accuracy ε, test sample stability threshold N_r, and the annealing ending threshold I, Calculate $f(X_0)$, make $X_{opt} = X_0$, $f_{opt} = f(X_0)$.

Step 2. Using simulated annealing algorithm to determine the weight of the combined model automatically. If the result does not meet the accuracy, use $Y_i = X_i + \eta\gamma V_0$ to generate a new dynamic point, in which, η is the control of disturbance ratio constant, γ is Cauchy random disturbance and V_0 is the step value.

Calculate $f(X_i)$.

If $f(Y_i) \le f(X_i)$ then $f(X_{i+1}) = f(Y_i)$, otherwise. Calculate $P = \exp[(f(X_i) - f(Y_i))/T_k]$, if $PP \le P$, then $X_{i+1} = Y_i$, $f(X_{i+1}) = f(Y_i)$, otherwise,

$X_{i+1} = X_i$, $f(X_{i+1}) = f(X_i)$.

Step 3. Update the weights

If $f(X_{i+1}) < f_{opt}$, then $X_{opt} = X_{i+1}$; $f_{opt} = f(X_{i+1})$, $m = 0$, otherwise, $m = m + 1$.

Step 4. MetroPolis' sampling Stability Criteria

If $m < N_r$, then $i = i + 1$, and turn to step 2; otherwise, $m = 0$.

Step 5. Annealing ending discrimination

If $f_{opt} > T_k$, then $i_k = i_k + 1$, otherwise, $i_k = 0$. If $i_k \ge I$ or $f_{opt} < \varepsilon$, annealing finishes, otherwise, go to step 6.

Step 6. Annealing method

If $f_{opt} > f^*_{opt}$, then $T_{k+1} = T_0 / 1 + delta \bullet (func\_call)$ to anneal, otherwise, according $T_{k+1} = r_T \bullet T_k$ to anneal, set $k = k + 1$, and go to step 2. Where $func\_call$ is the objective function using times and $0 < r_T < 1$.

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 Fig.1.

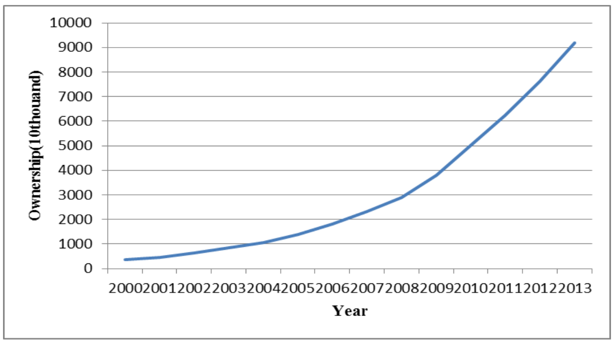

FIGURE 1 China's private car ownership from 2000 to 2013

TABLE 1 The actual value and predictive results.

Year	Actual value	Regression	GM(1,1)	SVM	Combined Model
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	844.456
2004	1,069.69	1141.258	1269.558	1067.519	1069.731
2005	1,383.93	1412.722	1602.157	1386.165	1386.962
2006	1,823.57	1793.855	2021.891	1829.026	1827.971
2007	2,316.91	2327.081	2551.587	2319.145	2319.383
2008	2,880.50	2925.439	3220.052	2885.269	2886.474
2009	3,808.33	3609.001	4063.643	3810.726	3804.674
2010	4,989.50	4734.339	5128.238	4987.238	4979.651
2011	6,237.46	6166.945	6471.735	6234.573	6232.544
2012	7,637.87	7680.559	8167.202	7634.73	7636.105
2013	9,198.23	9379.076	10306.85	9196.059	9201.55
MAPE	--	2.680%	10.761%	1.284%	1.252%

It can be observed that the curve has an increasing linear tendency. Therefore, the regression model is chosen as a constitution of the combined forecasting model. In addition, the models of grey forecasting model GM(1,1) and support vector machine (SVM) are pitched on in consideration of their popularity as form a combined model. In SVM 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. The forecasting results and the combined model with simulate anneal algorithm optimized weights are shown in Table 1.

Judging from the determined weight of simulate anneal algorithm, we can infer that SVM model gets the maximum weight, followed by regression. The GM(1,1) model's weight is the minimum. Generally speaking, the better the performance is, the greater the weight is. And the weight optimized by simulate anneal algorithm is also proven it; the performance of the proposed model is the best one.

4 Conclusions

The combined model applied in this paper is an integrated method, which is composed of regression model, GM(1,1) model and SVM model. By means of simulate anneal algorithm, a more reasonable weight coefficient can be obtained and the predictive effect, in allusion to college employment population, tends to be accurate. In the experiment, the proposed model gets the best performance which verifies its effectiveness on the aspect of predictive capability.

References

[1] Azadeh A. (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), 221-240.

[2] Wu J D, Liu J C. (2012) A forecasting system for car fuel consumption using a radial basis function neural network. *Expert Systems with Applications*, 39(2),1883-1888.

[3] Zheng X. (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*, 15(4), 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),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), 1210-1226.

[6] Pierdzioch C, Rulke, Stadtmann G. (2011) Forecasting US car sales and car registrations in Japan: Rationality, accuracy and herding. *Japan and the World Economy*, 23(4), 253-258.

[7] Lee D H. (2013) Analysis of the energy and environmental effects of green car deployment by an integrating energy system model with a forecasting model. *Applied Energy*, 103, 306-316.

[8] Qian L X, Soopramanien D. (2014) Using diffusion models to forecast market size in emerging markets with applications to the Chinese car market. *Journal of Business Research*, 67(6),1226-1232.

Authors

Junchao Yang, 12. 10. 1980, China

Junchao Yang received the management master degree in school of economic and management from North China Electric Power University, China in 2007. Currently, he is a doctor at North China and Electric Power University. His research interests electric vehicles project management.