Journal of Southeast University (English Edition) Vol. 29, No. 2, pp. 194 − 198 June 2013 ISSN 1003—7985

Research on development of urban taxi supply based on influence factors classification

Chen Jingxu Wang Wei Chen Xuewu Shen Jinshi

(School of Transportation, Southeast University, Nanjing 210096, China)

Abstract: In order to determine the regulations of the development of taxi supply under entry regulations in Chinese cities, an improved neural network model is applied to find the particular years when the government artificially puts new taxis into the market, and then extract the political influence from the taxi supply. The model is also utilized to study the relationships between the adjusted taxi supply and non-policy factors. A case study of Nanjing city is conducted. The results show that 2001 and 2007 are the particular years that the Nanjing government artificially put new taxis into its taxi market, which is in accordance with the five-year plan of China and the local development plans. The results also show that the improved neural network model has a good performance in expositing the evolution of adjusted taxi supply related to non-policy factors.

Key words: taxi supply; neural network model; policy year; influence factor

doi: 10.3969/j. issn. 1003 −7985. 2013. 02. 015

The taxi is gradually becoming a supplement to public transport due to its convenience, flexibility and door-to-door services. A number of studies have been conducted on different kinds of regulations with respect to the taxi market[1−4] and the development of taxi fleet size[5−7]. However, most of the current literature about the regulation of the taxi market comes from developed areas whose environments are quite different from those of mainland China. In China, entry regulation came into effect in the late 1990s. Despite limiting entry to the taxi industry, the government may artificially increase several hundred new taxis in one year as a result of the necessity of the urban development plans or for some occasional events (like holding an international sporting event). Such a particular year is described as a policy year for the taxi industry.

The development of taxi supply is also influenced by many non-policy factors, which can be divided into four parts, namely socioeconomic indices, scale and layout indices, other trip mode indices and tourism indices. But some newly-added number of taxis by the government as mentioned above cannot be explained by these factors. Thus, influence factors relevant to taxi supply involve two aspects: policy factor and non-policy factors (see Fig. 1).

This paper concentrates on the analysis of the relationship between the taxi supply and influence factors. It utilizes an improved back-propagation neural network model. The model attempts to find out the policy years of the taxi industry and then investigate the relationship between the taxi supply and non-policy factors. A case study of Nanjing city is performed.

1 Modeling

A back-propagation (BP) neural network has advantages of self-learning, self-adapting and robustness. A three-layered BP neural network can be used to simulate nonlinear functions[8]. However, it has a poor speed of constringency, and it is easy to get stuck in a locally optimal solution. In order to overcome such disadvantages, a method combining the genetic algorithm (GA) with a BP neural network is used[9]. As a heuristic stochastic search algorithm, the GA does well in global searching, which can optimize the weights and thresholds in the BP neural network. Based on the improved BP neural network, this paper focuses on the identification of policy years and the relationship between the taxi supply and the influence factors. The process of the model is described below.

1.1 Normalization of data

Fifteen non-policy factors are scaled to the same range between 0 and 1 as

$$\mu_{ij} = \frac{X_{ij} - \min(X_j)}{2(\max(X_j) - \min(X_j))} \tag{1}$$

where X_{ij} is the value of the j-th factor in the i-th year; $\max(X_j)$ is the maximum in the data set of the input sector X_j; and $\min(X_j)$ is the minimum in the data set of the input sector X_j.

The normalization of the taxi supply, which makes the value in a fixed interval [0, 1], is given by

$$n_i = \frac{N_i - N_0}{N_{\max} - N_0} \tag{2}$$

Received 2012-12-07.

Biographies: Chen Jingxu (1989—), male, graduate; Chen Xuewu (corresponding author), female, doctor, professor, chenxuewu@seu.edu.cn.

Foundation item: The National Basic Research Program of China (973 Program) (No. 2012CB725400).

Citation: Chen Jingxu, Wang Wei, Chen Xuewu, et al. Research on development of urban taxi supply based on influence factors classification [J]. Journal of Southeast University (English Edition), 2013, 29(2): 194 − 198. [doi: 10.3969/j. issn. 1003 −7985. 2013. 02. 015]

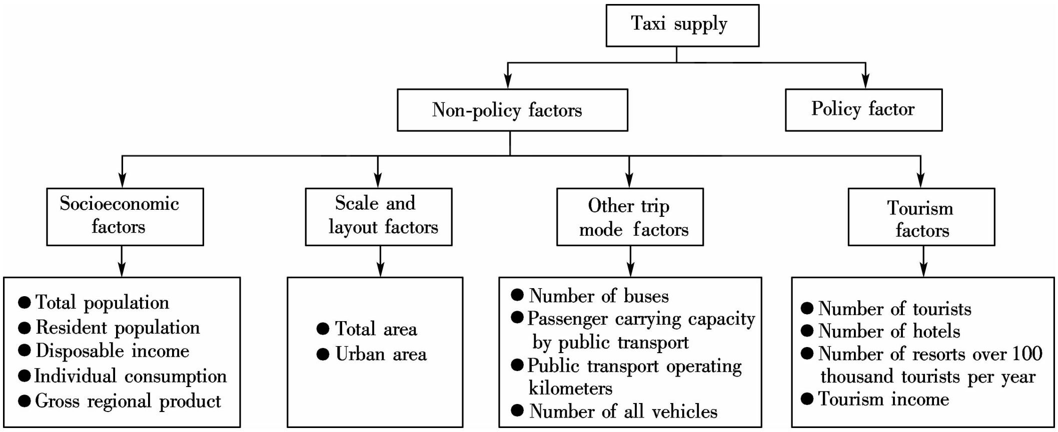

Fig. 1 Influence factors relevant to the taxi supply

where N_i is the taxi supply of the i-th year; N_0 is the taxi supply of one year when the taxi market is in a free entry condition; and $N_{\max}$ is the maximum taxi quantity based on the logistic model[10].

1.2 Model description

Fig. 2 depicts the parallel distributed network which consists of three parts: an input layer, one hidden layer and an output layer.

In the input layer, fifteen non-policy factors after normalization procedures $\{\mu_{i1}, \mu_{i2}, \dots, \mu_{i15}\}$ are put into the input layer. The input layer has 15 nodes. In the hidden layer, the k-th hidden node ν_{ik} is expressed as

$$\nu_{ik} = \tanh\left(\sum_{j=1}^{15} \mu_{ij} w_{ijk}^{(1)} + b_{ik}^{(1)}\right) \qquad k = 1, 2, \dots, p \quad (3)$$

where p is the number of hidden nodes; $w_{ijk}^{(1)}$ is the weight between the j-th node in the input layer and the k-th node in the hidden layer; $b_{ik}^{(1)}$ is the threshold value of the k-th-hidden node. Here the hyperbolic tangent function is selected as the nonlinear activation function. The number of hidden nodes is determined by the method of trial-and-error, and it is found that the best number is 31 ($p = 31$).

In the output layer, there is only one node $\hat{n}_i$. The output is expressed as

$$\hat{n}_i = \log\left(\sum_{k=1}^{p} \nu_{ik} w_{ik}^{(2)} + b_i^{(2)}\right) \qquad (4)$$

where $w_{ik}^{(2)}$ is the weight between the k-th node in the hidden layer and the node in the output layer; $b_i^{(2)}$ is the threshold value of the output neuron. Here the logistic sigmoid function is selected as the activation function.

The optimization of weights and thresholds based on the GA is as follows:

Step 1 According to the established neural network, we use the real number form to encode the initial population A by a complete set of parameters $w_i = \{w_i^{(1)}, w_i^{(2)}, b_i^{(1)}, b_i^{(2)}\}$. Here the GA is run with a population size of 50 and the number of evolution generations is set to be 200.

Step 2 Evaluate the fitness of all chromosomes by

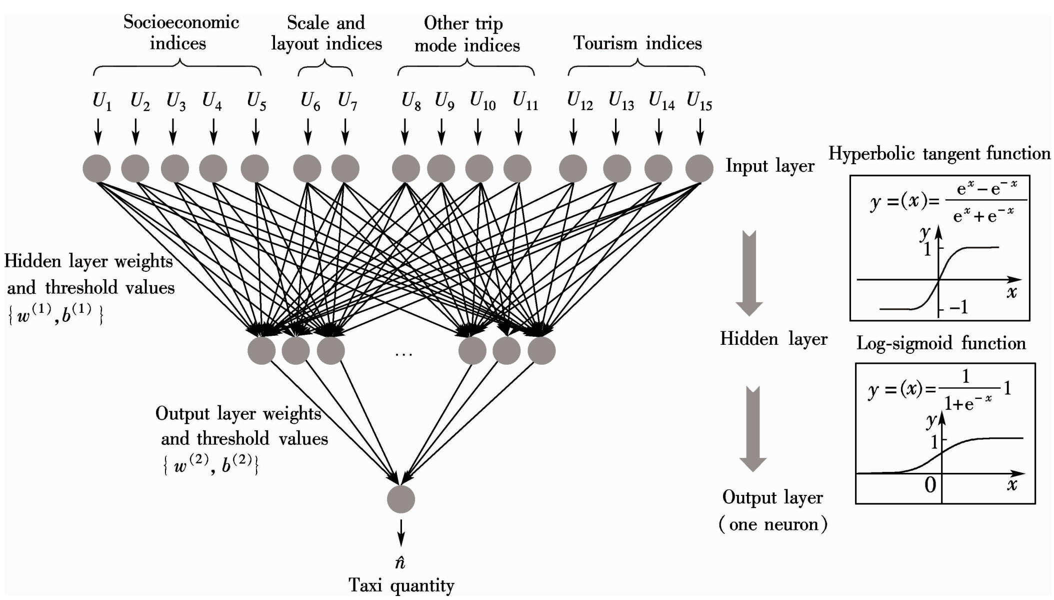

Fig. 2 Basic architecture of the neural network used in this study

constructing the corresponding neural network. The objective function is described as

$$f = \frac{1}{\sum_{i=1}^{N}(n_i - \hat{n}_i)^2} \tag{5}$$

where n_i is the desired output; $\hat{n}_i$ is the actual output; and N is the number of years in the training data set.

Step 3 Conduct the selection operation. We use the roulette wheel selection based on the ranking algorithm. After conducting the selection operator, the intermediate population is created by extracting chromosomes from the current population. The selection probability for the individual m is

$$p_s = \frac{f_m}{\sum_{m=1}^{M} f_m} \tag{6}$$

where M is the number of chromosomes and f_m is the fitness of individual m.

Step 4 Execute the crossover operation. According to the crossover probability p_c (Assume that $p_c = 0.7$), the parent individuals produce a new generation through linear crossovers. The chromosomes a_m and a_l conducting the crossover process at h is expressed as

$$\left.\begin{aligned} a_{mh} &= a_{mh}(1-b) + a_{lh}b \\ a_{lh} &= a_{lh}(1-b) + a_{mh}b \end{aligned}\right\} \tag{7}$$

where b is the random number located in the interval (0, 1).

Step 5 Implement the mutation operation. The mutation parent individuals are randomly chosen based on the mutation probability p_m (Assume that $p_m = 0.01$). The variation process for the gene a_{mh} is given as

$$a_{mh} = \begin{cases} a_{mh} + (a_{mh} - a_{\max})f(g) & r \geqslant 0.5 \\ a_{mh} + (a_{\min} - a_{\mathrm{mh}})f(g) & r < 0.5 \end{cases} \tag{8}$$

where $a_{\max}$ is the upper limit of a_{mh}; $a_{\min}$ is the lower limit of a_{mh}; r is the random number located in the interval (0, 1); $f(g) = r(1 - g/G_{\max})$; g is the current number of iterations; and $G_{\max}$ is the maximum number of generations.

Step 6 Examine whether the fitness degree exceeds the given precision requirement (It is set to be 10^8) or the number of generations attains the presumed maximum value. If it does not satisfy, repeat step 2 to step 5 until the terminal condition is met.

The learning process of the optimum BP neural network is as follows:

Step 1 The BP neural network weights and threshold values are initialized as the chromosome of the best fitness population member based on the GA.

Step 2 Calculate the values of the hidden layer and the output layer as presented in Eqs. (3) and (4).

Step 3 The mean square error function is adopted as the error function, which is defined as

$$E = \sum_{i=1}^{N} e_i = \frac{1}{2}\sum_{i=1}^{N}(n_i - \hat{n}_i)^2 \tag{9}$$

Step 4 The error in the output layer is propagated backward to hidden layer neurons, and then to input layer neurons revising the weights and threshold values by the Levenberg-Marquardt method.

Step 5 Repeat step 2 to step 4 until the error is reduced to a predetermined convergence tolerance or the iteration number attains the maximum iteration number (Assume that the target error is 10^{-6} and the maximum iteration number is 10^4).

2 Case Study

Nanjing is selected as the study area, and the data source in this paper comes from the statistical yearbook of Nanjing[11].

2.1 Identification of policy years of taxi industry

Nanjing began to implement entry regulations in 1997, and in the later three years there were no big events or exigent requirements for an expansion of the taxi market. Thus, it is assumed that there is no policy year of the taxi industry from 1997 to 1999. The identification procedure is as follows:

Step 1 Data from the year $t-3$, $t-2$, $t-1$ is put into the improved neural network model for training, and the data from the year t is used for identifying whether the year t is a policy year (The year 2000 is the first testing year). If the relative error between the desired output and the actual output exceeds 5%, the year t is identified as a policy year of the taxi industry and go to step 2; if not, go directly to step 3.

Step 2 The difference value between the desired output and the actual output can be seen as the newly-added taxi quantity in one policy year. As the local government may artificially put several hundred taxis one-off or in batches[3-4], the newly-added number is rectified as

$$\Delta N_t = \left\lfloor \frac{N_t^{\text{desired}} - N_t^{\text{actual}}}{50} \right\rfloor \times 50 \tag{10}$$

where N_t^{desired} is the desired output of the taxi quantity and N_t^{actual} is the actual output of the taxi quantity. After excluding these newly-added taxis, the remaining number of taxis is the adjusted taxi quantity from such a policy year to the last testing year 2011. And then go to step 3.

Step 3 Examine the year t. If the year t is the last testing year, 2011, the process of identification is over; if not, let $t = t + 1$ and return to step 1.

Tab. 1 shows the results of the identification of the policy years. The years 2001 and 2007 are the policy years of the taxi industry. Referring to the five-year plan of China and the local government reports about the taxi industry, it is noteworthy that the results of identification are basically in accordance with these development plans.

Tab. 1 Indentification of the policy years of taxi industry in Nanjing

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Taxi supply	8 597	8 956	8 877	9 216	9 098	9 055	9 262	9 997	10 151	10 364	10 593	10 644
Desired output	8 597	8 456	8 377	8 716	8 598	8 555	8 762	9 497	8 951	9 164	9393	9 444
Actual output	8788	8 453	8 308	8 395	8 772	8 655	8 525	8 754	8 859	9 010	9 304	9 496
Relative error/%	−2.22	5.62	0.82	3.68	−2.02	−1.17	2.70	7.82	1.03	1.68	0.95	−0.55
Artificially increasing quantity		500						700				
Adjusted taxi supply	8 597	8 456	8 377	8 716	8 598	8 555	8 762	8 797	8 951	9 164	9 393	9 444

* Note: Relative error $= \dfrac{\text{desired} - \text{actual}}{\text{desired}} \times 100\%$.

At the beginning of 2001, according to the 10th five-year plan (2001—2005), the local government of Nanjing decided to prioritize the tourism industry and encourage Nanjing to be an international tourism city. When tourists (especially foreigners) are in an unfamiliar tourist city, they may choose taxis as their main trip mode. Besides, the statistical yearbook of Nanjing indicates that the number of tourists ascended visibly during the 10th five-year plan. Therefore, it is reasonable that the local government put 500 new taxis into the taxi market in 2001, and 2001 should be a policy year. When it comes to the 11th five-year plan (2006—2010), the major plan of local government was primarily to develop two rural areas (Jiangning District and Pukou District), which are expected to become centers of technological research and innovative startups. But these two districts lag behind in transport infrastructure, and it may be a reason for the government to add 700 new taxis in 2007, and 2007 can be regarded as a policy year. The adjusted values of the taxi quantity, excluding the political influence, are presented in the last column of Tab. 1.

2.2 Interaction between taxi supply and non-policy factors

After subtracting the political influence, it is more accurate to get the interplay between the development of the adjusted taxi supply and the non-policy factors based on the established model once again.

The normalized data from 1997 to 2008 constitute the training set and the remaining normalized data (from 2009 to 2011) are used in the testing phase. The input vector is represented by fifteen normalized non-policy factors. Accordingly, the output vector represents the normalized values for the adjusted taxi quantity. The training parameters are as follows: The population size N is 50; the number of evolution generation is 200; the crossover probability p_c is 0.7; the mutation probability p_m is 0.01; and the training precision is 10^{-6}. The outputs and relative errors of the testing data are shown in Tab. 2.

Tab. 2 Simulation results and errors of testing data

Year	Desired output	Actual output	Error/%	Average error/%
2009	10 364	10 296	0.66	
2010	10 593	10 563	0.28	0.43
2011	10 644	10 682	0.36	

In Tab. 2, based on the trained neural network model, the actual output simulated with the testing input data shows good agreement with the adjusted taxi quantity. Compared with the original BP neural network and the ARMA time series forecasting method[12] (The average errors are 1.46% and 2.99%, respectively.), the model developed in this paper has a better performance in finding the relationship between adjusted taxi quantity and non-policy factors.

Besides, the model can be utilized to calculate the predicted values of 15 non-policy factors from 2012 to 2015, and then forecast the taxi supply of Nanjing. The predicted values of the taxi supply in the next four years (2012—2015) are 10 763, 10 842, 10 895 and 10 847, respectively. Overall, future taxi supply will continually experience a slow increase trend, while fluctuating within a narrow range. It is just in accordance with the characteristics of the taxi market in Nanjing under strict entry regulation.

3 Conclusions and Future Work

Since the late 1990s, the majority of Chinese cities have regulated their taxi markets. Meanwhile, the local government still artificially adds some new taxis for the necessity of the urban development plans. In this paper, an improved neural network model is utilized, and a case study of Nanjing city is performed. First, the model is applied to identify the policy years of the taxi industry and then adjust the taxi quantity to exclude the political influence. The identified results are in accordance with the five-year plan of China and the local government reports about the taxi market. Thus, 2001 and 2007 can be seen as policy years of the Nanjing taxi market. Then the interactions between the adjusted taxi supply and non-policy factors are studied based on the model once again. The simulation results exhibit that the model has a good performance in finding the relationship between the adjusted taxi supply and non-policy factors. In the future, we will establish a demand model of the taxi market and explore the demand-supply relationship in the taxi industry, and try to predict potential policy years of the taxi industry.

References

[1] Flores-Guri D. An economic analysis of regulated taxicab

markets [J]. *Review of Industrial Organization*, 2003, **23**(3):255－266.

[2] Schaller B. Entry controls in taxi regulation: implications of US and Canadian experience for taxi regulation and deregulation [J]. *Transport Policy*, 2007, **14**(6): 490－506.

[3] Yao Zhigang, Cheng Gao. Clustering taxi quantity regulatory levels of Chinese capital cities[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2012, **12**(5):1－6.

[4] Tian W J. Investigating the regulations of the taxicab market in Chinese large cities [D]. Wuhan: Zhongnan University of Economics and Law, 2010. (in Chinese)

[5] Xu J M, Wong S C, Yang H, et al. Modeling level of urban taxi services using neural network [J]. *Journal of Transportation Engineering*, 1999, **125**(3): 216－223.

[6] Yang H, Lau Y W, Wong S C, et al. A macroscopic taxi model for passenger demand, taxi utilization and level of services [J]. *Transportation,* 2000, **27**(3): 317－340.

[7] Salanova J M, Estrada M, Aifadopoulou G, et al. A review of the modeling of taxi services [J]. *Procedia—Social and Behavioral Sciences*, 2011, **20**: 150－161.

[8] Whitley D, Starkweather T, Bogart C. Genetic algorithms and neural networks: optimizing connections and connectivity [J]. *Parallel Computing*, 1990, **14**(3): 347－361.

[9] Wen X L, Song A G, Duan J H, et al. Evolving neural networks using an improved genetic algorithm [J]. *Journal of Southeast University*: *English Edition*, 2002, **18**(4):367－369.

[10] Miranda L C, Lima C A. On the logistic modeling and forecasting of evolutionary processes: application to human population dynamics [J]. *Technological Forecasting and Social Change*, 2010, **77**(5): 699－711.

[11] National Bureau of Statistics of China. *The statistical yearbook of Nanjing*[M]. Beijing: China Statistics Press, 1995—2012.

[12] Box G E P, Jenkins G M. *Time series analysis: forecasting and control* [M]. San Francisco, CA, USA: Holden Day, 1978.

基于影响因素分类的城市出租车保有量发展规律

陈景旭　王　炜　陈学武　沈劲石

(东南大学交通学院,南京 210096)

摘要:为研究我国城市出租车保有量在准入规制条件下的发展规律,应用改进的神经网络模型寻找政府人为投入出租车运力的年份,并将政策因素的影响从出租车保有量中分离.继而研究调整后的出租车保有量与非政策性影响因素间的关系.以南京市为例进行建模分析,结果表明2001年与2007年是南京市政府人为投放出租车运力的年份,与中国五年计划及当地政府规划一致.与此同时,实例所得出的结果显示改进的神经网络模型对揭示分离出政策影响后的出租车保有量随非政策性影响因素的发展规律有很高的精度.

关键词:出租车保有量;神经网络模型;政策年份;影响因素

中图分类号:U491