Applied Computer Letters (ACL)

ARTICLE

GENETIC NEURAL NETWORK PREDICTION OF CAR OWNERSHIP BASED ON PRINCIPAL COMPONENT ANALYSIS

Yewang Zhou

School of Business, Huanggang Normal University, Huanggang 438000, China
Corresponding Author Email: whuw@sina.com

ARTICLE DETAILS

ABSTRACT

The prediction of the car ownership is the basic work for city traffic sustainable development. The paper carry on principal component analysis to influencing factors in the process of prediction of Wuhan City car ownership, determine the main components. Combining genetic algorithm with neural network, using genetic algorithm to optimize the weights of neural network, determine the initial weight values of neural network. Not only to improve the neural network training speed and generalization ability, but also overcome that the network is easy to fall into local minimum to a certain extent, then train the neural network, and carry on the prediction of car ownership. At last, use a specific example to verify the prediction effect.

KER WORDS

1. INTRODUCTION

Study on Car ownership is a basic work of city road traffic planning, how to accurately predict the city car ownership is of great significance to the sustainable development of city traffic planning and management. The prediction of the car ownership is the basic work for city traffic sustainable development. Lam et al set the city road network capacity and parking facilities as two constraints of car ownership development, using the theory of bilevel programming to construct the car ownership prediction model [1][2]. Tomas et al established dynamic car ownership prediction model based on the 2D model, Rich established micro economics model on prediction of car ownership based on random utility theory and micro economics in the comprehensive consideration family and work site [4].

The current domestic car ownership prediction method can be divided into the following three categories: one is the extrapolation method, namely using past data to predict the future state, such as time series analysis. The second method is the law of cause and effect, according to the historical data, to find the relation between the variables to be forecasted and related variables, so as to predict the future state, as the regression analysis method. The third method is judgment analysis method, which mainly depends on the expert experience and comprehensive analysis ability to predict future states of [5][6].

Because of the diversity and complexity of traffic system influencing factors, although the current prediction methods have their advantages, but the data fit is not high, the prediction results have deviation from the actual. Such as time sequence prediction method, using historical data of car ownership to predict the time trend, which does not take relative influencing factors into account. Regression analysis method establish

regression model based on some influencing factors that influence the amount of car ownership, but it only consider several influencing factors, Simulation ability to practical problems is not accurate. With the exception of the above methods, domestic scholars are still using some new artificial intelligence method to predict the car ownership [7].

Genetic algorithm is a nonlinear optimization method. Combining genetic algorithm with neural network, we can have both extensive mapping ability of neural network and rapid convergence properties of genetic algorithm. The paper combine genetic algorithm with neural network, using genetic algorithm to optimize the weights of neural network, determine the initial weight values of neural network. Not only to improve the neural network training speed and generalization ability, but also overcome that the network is easy to fall into local minimum to a certain extent.

2. PRINCIPAL COMPONENT ANALYSIS OF CAR OWNERSHIP

2.1 Basic Theory of Principal Component Analysis

Principal component analysis (Principal Component Analysis, PCA) is a multi variable statistics analysis technique for data compression and feature extraction, can effectively remove the correlation among data. Its purpose is to guarantee the data loss as small as possible, through linear transform abandon small part of the information, use a few new variables (called principal components) to replace the original multidimensional variable, requirements of the main components can fully reflect the information of the original variables, at the same time each principal component is uncorrelated.

Using the extracted principal components as the training data of neural

network, not only passed the original variable information, but also reduces the number of variables, reduces the complexity of the neural network, can improve the performance of neural network as a whole [8].

2.2 Principal Component Analysis of Wuhan City Car Ownership

According to influencing factor analysis of car ownership and the feasibility of data, this paper selected some index from 1995 to 2012 at the end of Wuhan city population (person), gross domestic product (100 million yuan), the total industrial output value (100 million yuan), the total import and export volume (million), the total retail sales of consumer goods (100 million yuan), fuel dynamic price index (with a base year of 1995 100), urban residents' per capita disposable income (yuan). The total social investment in fixed assets (100 million yuan), total social freight volume (million tons), the whole society of passenger volume (million), pavement, road length (km) and eleven indexes to predict Wuhan city car ownership quantity. These eleven indexes can be divided into population index, economic index, road traffic index.

In this paper X1, X2...X11 represent 11 index, using the actual data and principal component analysis can determine the three principal component expressions are as follows:

$$
\begin{aligned}
Y1 = &0.3005X1 + (-0.4626)X2 + (-0.1068)X3 + (-0.0132)X4 \\
&+ 0.4888X5 + (-0.1938)X6 + (-0.3654)X7 + (-0.2702)X8 \\
&+ 0.1636X9 + 0.3898X10 + 0.0733X11
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
Y2 = &-0.4626X1 + (-0.1068)X2 + (-0.0132)X3 + 0.4888X4 \\
&+ (-0.1938)X5 + (-0.3654)X6 + (-0.2702)X7 + 0.1636X8 \\
&+ 0.3898X9 + 0.0733X10 + 0.3309X11
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
Y3 = &0.1535X1 + 0.0251X2 + (-0.082)X3 + (-0.5382)X4 \\
&+ 0.138X5 + (-0.2959)X6 + (-0.05)X7 + (-0.2688)X8 \\
&+ 0.6625X9 + 0.2341X10 + 0.0284X11
\end{aligned} \tag{3}
$$

The cumulative variance contribution ratio of the three principal components reached 98.844% can fully reflect the information of the original data.

3. CONSTRUCT PREDICTION MODEL

In this paper, using the global search ability of genetic algorithm to determine the initial weight values of neural network, in instead of random determination method of the initial weights of neural network, the method can solve the slow convergence speed of neural network, easily fall into local minima [9][10].

In this paper, combine the genetic algorithm with neural network to predict the Wuhan city car ownership, the basic idea is as follows:

(1) Principal component analysis of influencing factors, to determine the principal component;

(2) Determine the initial weights of neural network with the genetic algorithm;

(3) To train the neural network, and carry on the prediction of the car ownership.

3.1 Determine Neural Network Structure

1) Determine the number of neurons

According to the results of PCA, determine three principal components as input data, neuron number of input layer is 3. The output layer is the data of car ownership, neuron number of output layer is 1. This paper identify neurons in hidden layer is 15, thus the structure of neural network is 3*15*1.

2) Determine the training and learning function

Aiming at the shortcoming of the traditional BP algorithm, this paper selects the adaptive learning rate traingda as the training function of neural network, learngdm as the learning function of neural network.

3.2 Determine genetic Algorithm Structure

In this paper, using GAOT toolbox (Genetic Algorithm Optimization Toolbox) to achieve the genetic algorithm.

1) Determine coding mode

This paper select a floating-point number as a method of coding, the individual code string length: S=M * N + N * L + N + L. The basic form of coding is shown as follows:

$$ W = \{w_1, w_2, b_1, b_2\} \tag{4} $$

w_1 is the connection weights from the input layer to the hidden layer, the length is M * N;

w_2 is the connection weights from output layer to the hidden layer, the length is N * L;

b_1 is threshold value of the hidden layer, the length is N; b_2 is threshold value of the output layer, the length is L;

M, N, L are neuron numbers of the input layer and hidden layer.

2) Determine fitness function

This paper selects the performance function of neural network as the fitness function of genetic algorithm.

Suppose neural network has n learning samples, then:

$$ E = \sum_{i=1}^{n} (T_i' - T_i)^2 \tag{5} $$

$$ f = \frac{1}{E} \tag{6} $$

E is neural network performance function, T' is the actual output of the network, T is the expected output of the network, F is the fitness function of genetic algorithm.

3) Determine the selection operator

The paper chooses selection probability of operator as q, take the default value is 0.08. The paper chooses arithmetic crossover as the operator of crossover algorithm, and the crossover probability is 0.95.

The paper chooses the non-uniform mutation as the mutation operator, the mutation probability is determined as 0.08. The paper chooses the maximum genetic algebra as the termination condition of algorithm, set the maximum genetic algebra as 200 generation.

4. PREDICTION ANALYSIS

4.1 Genetic Neural Network Prediction Based on Principal Component Analysis

Based on the genetic algorithm and the neural network structure, using the matlab7.0 to program and select data from 2000 to 2014 as training data, select data from 2015 to 2017as test data of the network prediction performance.

1) The genetic algorithm optimization

By using genetic algorithm for global optimization, get an initial weights, threshold value matrix of neural network, the next is evolutionary graph of error sum of squares.

Error sum of squares show the performance of neural network, solve reciprocal to error sum of squares and get the fitness in the genetic evolution process, as shown below:
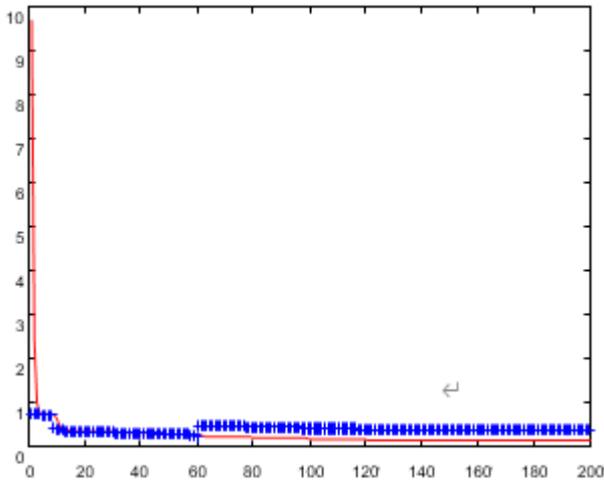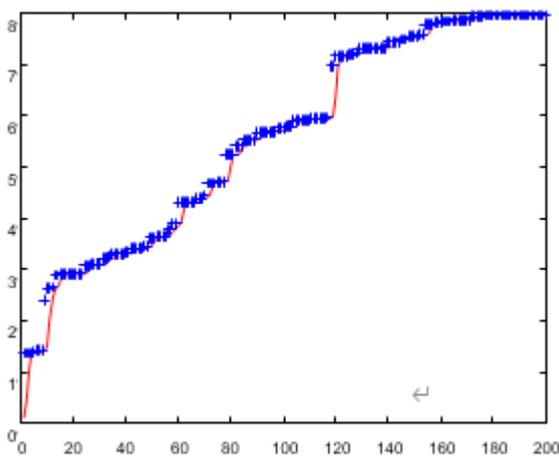
Figure 1: Evolutionary graph of error sum of squares

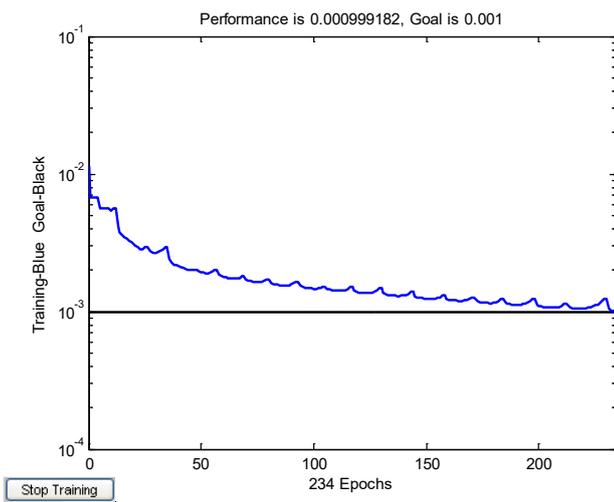**Figure 2:** The genetic evolutionary fitness map

Figure 3: BP network training

When in the evolution to the 160 generation, fitness gradually become stable, when in the evolution to 200 generation it was obtained the initial weights, threshold value matrix of neural network.

2) Train neural network and test prediction

Put the initial weights and threshold matrix be optimized by the genetic algorithm in neural network, and set the network training accuracy as

10^{-3}, neural network training process is shown below:

Through 234 times training, the neural network begins to convergence, the process spends 19.9 seconds. Use the trained network to predict 2015, 2016, 2017 in Wuhan city car ownership, can get prediction results:

Table 1: Prediction performance table

Type	2015	2016	2017
Practical value	370609	418667	484111
Prediction value	380653	424654	485805
Absolute error	2.71	1.43	0.35
The mean absolute error		1.49	

As in 2015 car ownership growth rate dropped from 23.2% in 2014 to 10.8%, change is larger, the prediction result is too large. Prediction results mean absolute error is 1.49%, meet the requirements of precision.

4.2 Compare With Prediction Method Based On Principal Component Neural Network

In order to be able to compare with prediction method based on principal component neural network, here keep network input, network structure and the other parameters unchanged, only delete optimization process of genetic algorithms, you can create BP network based on principal component analysis, carry on predicting to Wuhan City car ownership. BP network training process is shown below.
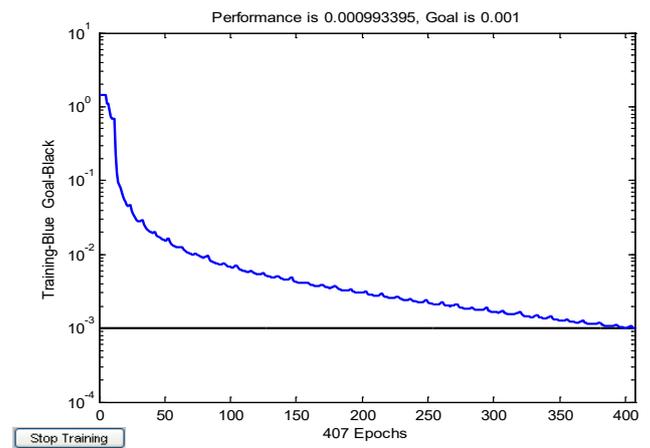

Figure 4: The traditional neural network training

The prediction results are shown in the following Table 2:

Table 2: Prediction performance comparison

Type	2015	2016	2017
Practical value	370609	418667	484111
BP prediction	395551	422644	461939
Error (%)	6.73	0.95	-4.58
PCA-GA-BP prediction	380653	424654	485805
Error (%)	2.71	1.43	0.35

Although the prediction error of traditional BP is smaller in 2016, but the mean absolute error of traditional BP is 4.1%, the prediction error is larger than PCA-GA-BP, and because the network dimension of the traditional BP is large, it spend 64.5 seconds on beginning to convergence. Therefore, through the principal component analysis of influencing factor, reducing the complexity of the network, but also use global optimization of genetic to determine the initial weight values of

neural network, combining with local optimization of the neural network, it can improve the prediction accuracy, but also reduce the complexity of network, which can satisfy the prediction demand.

5. CONCLUSION

Genetic algorithm is a nonlinear optimization method. Combining genetic algorithm with neural network, we can have both extensive mapping ability of neural network and rapid convergence properties of genetic algorithm. The paper combines genetic algorithm with neural network, using genetic algorithm to optimize the weights of neural network, determine the initial weight values of neural network. Not only is to improve the neural network training speed and generalization ability, but also overcome that the network easily falls into local minimum to a certain extent. The paper takes Wuhan City car ownership as the research object, and an example show that the method can satisfy the prediction demand.

REFERENCE

[1] Mei L.T., William, H.K. Lam., 2004. Balance Of Car Ownership Under User Demand and Road Network Supply Conditions — Case Study in Hong Kong. Urban Planning & Development, 24-36.

[2] Tam, M.L., William, H.K. Lam., 2000.Maximum Car Ownership Under Constraints of Road Capacity and Parking Space. Transportation Research Part.

[3] Tomas, S., 2004. Dynamical Models of Car Ownership at The Household Level. Working Paper.

[4] Rich, J.H., Nielsen, O.A., 2001. A Microeconomic Model for Car Ownership, Residence and Work Location. Paper for European Transport Conference PTRC，Cambridge.

[5] Ma, Y.L., 2007. Study On Prediction of Our Country Car Ownership. Traffic Science & Technology.

[6] Gong, H.W., 2005. Prediction of Car Ownership Based on Econometric Models. Journal Of Transportation Systems Engineering and Information.

[7] Hua, Z.G., 2005. Prediction of Hunan Province Car Ownership Based on BP Artificial Neural Network. Shanxi Science and Technology.

[8] He, M., 2007. Application Of Principal Component Analysis and BP Neural Network Method in The Prediction of Car Ownership. Computer and Communications.

[9] Cheng, C.S., Su, L.L., 2005. Prediction of Car Ownership in China Based on Genetic Algorithm BP. Journal of Changsha Institute traffic.

[10] Zhu, B., He, C.S., 2006. Prediction of Sichuan Civilian Vehicle Ownership Based on G M D H Method. Modern Management Science.