

Jun Hai-min – Pei Yu-long \*

## NEURAL NETWORK OPTIMIZED BY GENETIC ALGORITHM OF MODELS FOR REAL-TIME FORECAST OF TRAFFIC FLOW

*Short-term traffic flow forecast plays an important role in transit scheduling. A high-order generalized neural network model is constructed to actualize dynamic forecast on-line and a hybrid genetic algorithm and identical dimension recurrence idea are performed to optimize the structure and shape of neural network dynamically so as to enhance its forecast accuracy. With data collected from Dazhi Str., Harbin as the system input, the experimental result indicates that the average relative error of forecast is 5.53% and the maximum is less than 21%, which proves that the proposed neural network model can satisfy the precision request, accelerate the convergence speed, improve the global generalization ability and possess the practicality in short-term traffic flow forecast.*

*Key words – traffic flow forecast; neural network; genetic algorithm; relative error*

### 1. Introduction

Real-time dynamic traffic assignment is the main theory base in Intelligent Transportation System, while real-time traffic volume predicting is the precondition of dynamic traffic assignment. The result of traffic volume predicting is relative to traffic assignment, so how to predict traffic volume precisely [1-3] is the starting point.

By far, many statistical methods, i.e. multiple linear regression, stochastic time series, general exponential smoothing etc. have been used for traffic flow forecast. However, as a joint-effect result of weather and traffic environment, etc, its transfer regularity always reveals the representative complexities and nonlinear characters and fails to yield satisfied results [4]. For the past few years, artificial neural network attracts intensive attention and is proposed as a powerful computational tool to solve the complex problem, such as short-term traffic flow forecast.

Traditional forecasting techniques cannot generate accurate results under an environment with high complexity, randomness and non-determinacy, because they don't have self-adaptive and self-learning functions [5, 6].

Recent wavelet neural network, as another fruitful technique combining the merits of neural networks with wavelet analysis, has been widely used in forecast [7]. Since wavelet behaves excellent performance in non stationary signal analysis and non linear function modeling, the wavelet-based neural network can provide much higher availability of convergence for approximation than ordinary multi-layer networks.

### 2. Traffic Flow Prediction Model

#### 2.1 Model Choice

Traffic volume shows the complexity at intersections, because of traffic flow itself and the structure of road network [8]. Owing to high complex and variety with time, it is hard to get precise equations and label it. So we aim to get forecasting models using neural network technique and adaptive forecast traffic volume.

An artificial neural network, consisting of plenty of neurons whose broad links each other, is a nonlinear dynamic system. It is generated basing on the study of organize configurations and behavior characteristics in biologic neural system. It has been testified theoretically: a three layer neural network can achieve all kinds of complicated non-linear maps. So we found models and predicted them adopting a three layer neural network which has a certain hidden layer according to relativity of traffic flow in an intersection.

#### 2.2 Structure of Neural Network

As the traffic flow has complex time varying and non-linear characteristics, we found a prediction model to satisfy the need for a real-time dynamic traffic assignment adopting the high-order neural network, as shown in Fig. 1.

The network training process is as follows

- 1) The forward calculation process calculates the output state of every layer neuron, at last we get the real output of the network.

\* Jun Hai-min<sup>1,2</sup>, Pei Yu-long<sup>1</sup>

<sup>1</sup> School of Transportation Science and Engineering, Harbin Institute of Technology, 73 Huanghe Rd., Harbin 150090, China, E-mail: junhaimin@163.com

<sup>2</sup> Dalian Urban Planning & Design Institute, 186 Changchun Rd., Dalian 116011, China

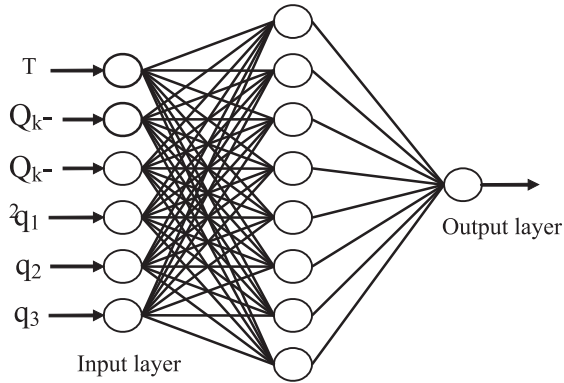


Fig. 1 Structure of neural network for traffic flow forecast

Here, we have get the optimal network nodes, that is 8, we choose a as 0.7 [9].

Suppose  $w_{ij}^{(h)}$ ,  $w_{jk}^{(o)}$ ,  $f_j^{(h)}$ ,  $f_k^{(o)}$  are the weights from the input layer to the hidden layer and from the hidden layer to the output layer and the adaptable Sigmoid parameter of the hidden layer and output layer neurons; then the input of hidden layer neurons

can be expressed as  $u_j^{(h)} = \sum_{i=1}^n w_{ij}^{(h)} \cdot o_i$  and the output of hidden layer neurons is given as  $o_j = f_j^{(h)}[u_j^{(h)}]$ .

The input of output layer neurons is presented by  $u_k^{(o)} = \sum_{j=1}^n w_{jk}^{(o)} \cdot o_j$  and the output of output layer neurons can be calculated through  $y_k = f_k^{(o)}[u_k^{(o)}]$ .

- 2) learning process according to the flowing equations, we adjust the network weight

Here the node state error and node reverse transfer error of output layer neurons can be obtained as  $\epsilon_k^{(o)} = t_k - y_k$  and  $\delta_k^{(o)} = \epsilon_k^{(o)} \cdot f_k^{n(o)}(u_k^{(o)})$ .

For hidden layer neurons, the node state error and node reverse transfer error are defined as

$$\epsilon_j^{(h)} = \sum_{k=1}^m \delta_k^{(o)} \cdot w_{jk}^{(o)} \quad (1)$$

$$\delta_j^{(h)} = \epsilon_j^{(h)} \cdot f_j^{n(h)}(u_j^{(h)}) \quad (2)$$

In the same meaning, we resume  $\eta$  is learning ratio,  $P$  is style-book gross and define  $\Delta w_{ij}^{(h)}$  and  $\Delta w_{jk}^{(o)}$  are separately the adjust quantity of weight from input layer to hidden layer and from concealed to output layer, which can be expressed as

$$\Delta w_{ij}^{(h)} = \eta \cdot \sum_{p=1}^P (\delta_{pj}^{(h)} \cdot x_{pi}) \text{ and } \Delta w_{jk}^{(o)} = \eta \cdot \sum_{p=1}^P (\delta_{pk}^{(o)} \cdot o_{pj}).$$

The following equation is the adjusted quantity of adapted parameter in the learning process:

$$\Delta a = \beta \cdot \epsilon \cdot x \quad (3)$$

where  $\epsilon$  is node state error,  $\beta$  is adjusted ratio of parameter,  $\Delta a$  is adjusted quantity of parameter.

In the training process, one of the key works depends on how to obtain the network weigh quickly and effectively. Upon this requirement, we induce genetic algorithm (GA) to optimize the weigh value as the following section.

### 2.3 Weigh Optimization by GA

The next problem is to conform hidden layer nodes. If there is no sufficient hidden layer nodes, it won't converge in learning process; if the number of hidden layer nodes is excessive, the network performance declines. The traditional BP algorithm conforms the number of hidden layer neurons on the basis of experience or examination; here we adopt a GA model to optimize the number of hidden layer neurons [10].

The GA approach, which is based on analogy with a natural selection and population genetics, is a kind of search method that was developed by John Holland. One normal application of GA is for searching an approximate solution to difficult optimization problems. However, local minimum value if the optimization function has a complex structure. In this paper, one of the main works is to find the optimization solution with a hybrid method, a genetic algorithm combined with a local heuristic search to offset the deficiency of the conventional genetic algorithm, for the network weigh [11].

Network parameters  $\beta$ ,  $w^h$  and  $w^o$  can be reached by minimizing an energy function, which is performed by the least-mean squares theory for data series representation.

$$\text{Min}E(\beta, w^h, w^o) = \frac{1}{2} \sum_{j=1}^n \sum_{i=1}^m [\tilde{y}_{ij}^k - y_{ij}^k]^2 \quad (4)$$

where  $\tilde{y}_{ij}^k$  and  $y_{ij}^k$  are the  $j$ -th desired and real obtained outputs with respect to the  $i$ -th input data.

The hybrid genetic algorithm involves initialization, mutation, competition and selection. An initial population is selected to initialize the parent generation according to a uniform distribution. In initialization, encoding takes on significant task and each population expressed by encoded chromosome stands for a certain solution to a required question.

A multi-valued encoding pattern is used to obtain a solution. A chromosome is divided into four segments: three of them are parameter segments, which are encoded in real values to deliver the information about the translation factor, telescopic factor and weight factor of the hidden unit. Simultaneously, the fourth segment is to encode the structure character in binary, where 1 represents a valid hidden unit, and 0 shows it is ineffective.

Each evolution generation is assigned a fitness value in accordance with the given fitness function

$$F(\beta, w^h, w^o) = \frac{1}{\sqrt{[\bar{y}_k^{ij} - y_k^{ij}]^2}} \quad (5)$$

since the most commonly used objective function aims to obtain the least error series between statistical data and forecast value.

Offspring is created from their parents by a composite mode. An encoding pattern of cross-gene is firstly judged and linear combination is used in a crossover operation for cross-gene encoded by real value. If  $x_i^g$  and  $x_j^g$  are defined as respective parent individual, the corresponding offspring through crossover is formulated by

$$x_i^{g+1}(P) = \lambda \cdot x_i^g(P) + (1 - \lambda) \cdot x_j^g(P) \quad (6)$$

$$x_j^{g+1}(P) = \lambda \cdot x_j^g(P) + (1 - \lambda) \cdot x_i^g(P) \quad (7)$$

where  $\lambda$  is an evolution rate complying with  $\beta^0 = 0.9$

Here,  $g$  and  $p$  separately represent a generation number of evolution and position of crossover gene. However, a conventional single-point crossover is preferred to further evolution, if crossover gene is encoded by binary.

Mutation is also treated with a different encoding mode and the mutation position encoded by real value yields the following mutation model

$$x_i^{g+1}(P) = x_i^g(P) + \delta \cdot E(x_{max}) \cdot P_m \quad (8)$$

where  $\delta$  is a stochastic value ranging over  $[-0.5, 0.5]$  and represents the cumulative value of a squared error derived from the chromosome with maximum fitness when evolving into the  $g$ -th generation and  $P_m$  is a mutation factor to control the mutating operation.

If the binary encoding is applied, offspring is generated from their parents by a Gaussian perturbation

$$x_i^{g+1}(P) = \lambda \cdot x_i^g(P) + N(0,1) \cdot P_{pi} P_m \quad (9)$$

where  $N(0,1)$  denotes a normally distributed one dimensional random variable with mean zero and standard deviation. To a total population pop,  $P_{si}$  is the  $i$ -th proportion factor

$$P_{pi} = \frac{F_{max} - F_i}{\sum_{i=1}^{pop} F_i} \quad (10)$$

The following parts (contents/factors) should be noted above all. The gradient descending factor, which holds the same iterative step-size with the evolution process, optimizes the network parameters without remodeling the structure of WNN and characteristic gene of optimized chromosome has to answer the limited scope, such as dilation factor. The next evolution group, conserving diversity of population through reproduction and competition, consists only of filial generation derived from crossover and muta-

tion, due to inheriting the unparalleled quality from paternal generation.

In the hybrid genetic algorithm, competition and selection lead to more adaptive generation through a given proportion, which inevitably causes premature or slow convergence. Here, the evolving principle is executed through optimum reservation strategy combined with proportion selection. All population is sequenced with ascending sort on the basis of fitness, among which chromosomes with large fitness will be assigned directly into the next generation. Otherwise, the above proposed evolving procedure has to be reduplicated via a selection factor

$$P_{si} = \frac{F_{max} - F_i}{\sum_{i=1}^{pop} F_i} \quad (11)$$

The selection factor assures the chromosome with low fitness to gain large probability in the next generation. However, evolving population is sometimes logging into a certain individual or such finite ones, which is usually named local convergence and reduces the probability of global optimum. Therefore, the distance between two individuals is significantly taken into account in selection operation and one individual has to be eliminated correspondingly due to a shorter distance.

### 3. Case Study

The following is an example of the optimized neural network framework using the real traffic volume data of the intersection of West Dazhi Str. in Harbin in Oct. 12~18, 2006. We used one-week traffic volume data to train, and Figs. 3, 4 are error variety curves with the increase of genetic step and variety process of hidden layer nodes with the increase of genetic step, respectively.

To determine a proper  $\beta$ ,  $w^h$  and  $w^o$ , the following simple scheme depends on proposed GA approach mentioned in section 2.3.

- 1) Pre-processing the input data series through fuzzy reasoning theory.
- 2) Generating the original population within a logical interval.
- 3) Utilizing crossover and mutation operations.
- 4) Creating new offspring via optimum reservation and proportion selection.
- 5) Sequencing the individuals by fitness and computing the distances between chromosomes.
- 6) Finishing the search process, if it satisfies the stopping rule, or going back to step III.
- 7) Real-time and dynamic forecast with identical dimension recurrence.

We train the neural network also using above traffic volume data when the objective is close to extreme. If we increase the learning ratio, the training time will decrease. So we design a variable learning ratio based on a genetic step and grads value, accelerating convergence speed. The training process with a proposed high-order generalized neural network; the training time is 894.02s,

that is about 15min, after that we get the network weight  $w_{ij}$  and  $w_{jk}$  by adaptable GA consideration, as from Fig. 2 to Fig. 5.

From the prediction result it can be seen that the average absolute error is 41, the maximum absolute error is 182; it is at the time interval 27. Figs. 4, 5 show the forecast error and average relative error is 5.53%, and thus we reach the maximum relative error 20.63%, at the time interval 19.

### 3. Conclusions

A considerable amount of experiences were gained through developing an improved neural network model to forecast the short term traffic volume in real-time traffic control. The experience leads to the following recommendations.

- 1) A multiple layer neural network model with learning function is discussed, for short term traffic flow prediction, and its struc-

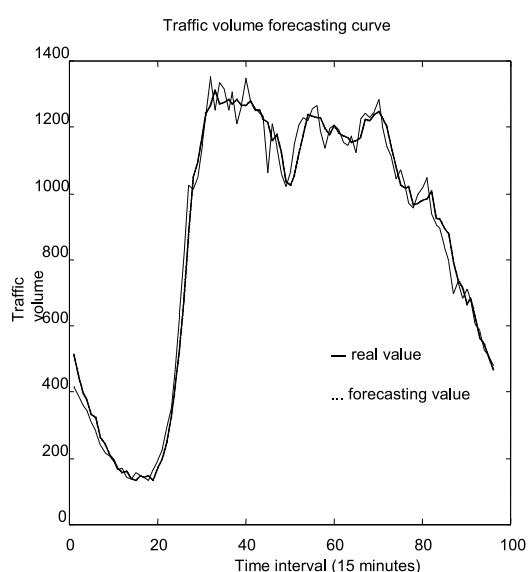


Fig. 2 Statistical and forecast volume of traffic flow

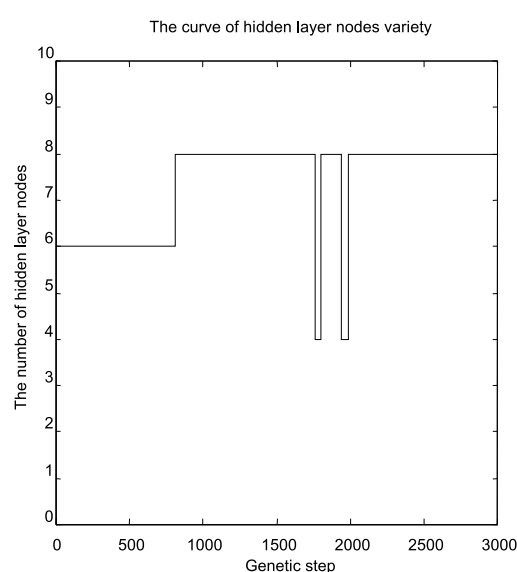


Fig. 3 Variety of hidden layer during training process

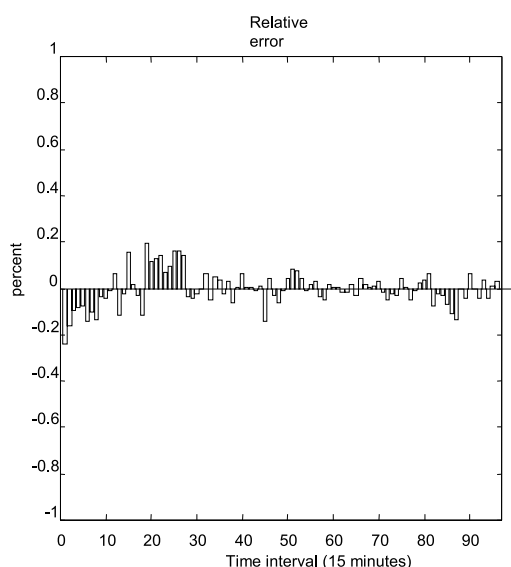


Fig. 4 Verification of relative error of traffic flow forecast

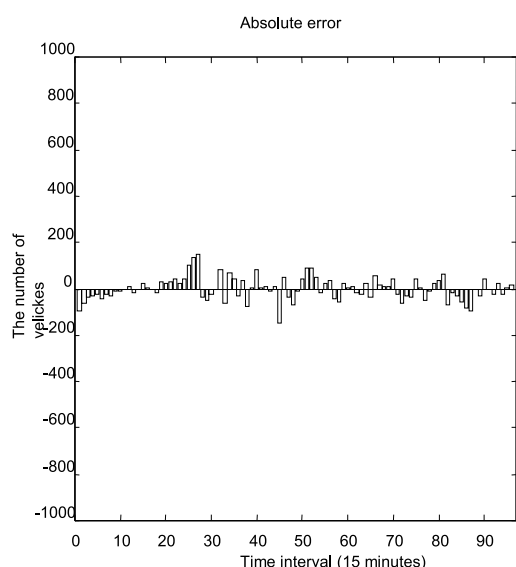


Fig. 5 Verification of relative error of traffic flow forecast

ture and working effectiveness depend on the optimization of three parameters, the weigh  $w_h$  and  $w_o$ , and the adjustment coefficient.

- 2) Hybrid genetic algorithm is introduced to optimize the parameters and structure of constructed neural network model. The experimental result indicates that the proposed neural network model optimized by the hybrid GA has fast convergence and high-accuracy ability.
- 3) Since traffic volume exists in the dynamic system of traffic and transportation, we can precisely forecast the traffic volume within several seconds, thus calculate travel time or obstruct function of a vehicle based on the traffic volume, ultimately plan the shortest route of every vehicle, in other words, guide the vehicle to move along an optimal route. This will significantly enhance efficiency of the whole traffic system, i.e. reduce traffic jams, save travel time, etc.

- 4) For future research it would be of interest to investigate factors affecting traffic flow and more optimization algorithms for real-time forecast to construct a practical operation system for dynamic traffic control and management.

#### Acknowledgment

This work is partially supported by the Key Projects in the National Science & Technology Pillar Program during the Eleventh Five-Year Plan Period of China, under granted number 2006BAJ18B03. Here we would like to extend our sincere gratitude to the editors, reviewers, and my colleagues in Institute of Traffic Engineering, HIT, for their fund sustentation, instructive advice and useful suggestions. Without their consistent and illuminating instructions, this research work could not have reached its present form.

#### References

- [1] EBERLEIN, X. J., WILSON, N. H. M., BARNHART, C., et al: *Real-time Deadheading Problem in Transit Operations Control*, Transp. Res. B Methodol., Vol. 32B, No. 2, pp. 77–100, 1998.
- [2] SITE, P. D., FILIPPI, F.: *Service Optimization for Bus Corridors with Short-turn Strategies and Variable Vehicle Size*, Transp. Res. A Policy Pract., Vol. 32, No. 1, pp. 19–38, 1998.
- [3] Kai, C., Mo, Z., Hamamatsu, Y.: *Traffic Flow Forecasting Using the DESFM Model*, IEEJ Trans. Electron. Inf. Syst., Vol. 126, No. 2, pp. 255–263, 2006.
- [4] LEDOUX, C.: *An Urban Traffic Flow Model Integrating Neural Networks*, Transport. Res. C Emerging. Technol., Vol. 5, No. 5, pp. 287–300, 1997.
- [5] WANG, Q., SHAO, H. H., ZHANG, Z. J.: *Genetic Evolved Neural Network and Its Application in Formaldehyde Process Modeling and Optimization*, J. Shanghai Jiaotong Univ. Sci., Vol. 30, No. 4, pp. 143–150, 1996.
- [6] HU, R. M., LI, D. R., XU, Z. Q., et al: *Generalized Information Storing Principle and Higher-order Generalized Neural Networks*, Acta Electronica Sinica, Vol. 24, No. 7, pp. 59–65, 1996.
- [7] XIANG, G., DONG, D.: *Activation Functions and a Method to Improve the Training of Neural Networks by Using BP Algorithm*, J. of Computer Research & Development, Vol. 34, No. 2, pp. 113–117, 1997.
- [8] ZHANG, J., WALTER, G. G., MIAO, Y. B. et al: *Wavelet Neural Network for Function Learning*, IEEE Trans. Signal Process., Vol. 43, No. 6, pp. 1485–1497, 1995.
- [9] FENG, J. N., DIAO, Z. J., YANG, X. B., et al: *Wavelet Neural Network Optimization Based on Hybrid Hierarchy Genetic Algorithm*, Syst. Eng. Electron., Vol. 27, No. 6, pp. 1112–1114, 2005.
- [10] KIM, K. H., PARK, J. K., HWANG, K. J., et al: *Implementation of Hybrid Short-term Forecasting System Using Artificial Neural Networks and Fuzzy Expert Systems*, IEEE Trans. Power Syst., Vol. 10, No. 3, pp. 1534–1539, 1995.
- [11] PEI, Y. L., WANG, Y. G.: *Real-time Optimization Algorithm with Evolving Fuzzy Wavelet Neural Network for Passenger Flow Prediction in Dynamic Transit Scheduling*, Dynamic of Continuous, Discrete, and Impulsive Systems, Series A Mathematical Analysis, Vol. 13, No. c, pp. 666–670, 2006.