

## Research Article

# Forecasting Beijing Transportation Hub Areas's Pedestrian Flow Using Modular Neural Network

Shuwei Wang,<sup>1</sup> Ronggui Zhou,<sup>1</sup> and Lin Zhao<sup>2</sup>

<sup>1</sup>Research Institute of Highway, Ministry of Transport, No. 8 Xi Tu Cheng Road, Haidian District, Beijing 100088, China

<sup>2</sup>Key Laboratory of Transportation Engineering, Beijing University of Technology, No. 100 Ping Le Yuan, Chaoyang District, Beijing 100124, China

Correspondence should be addressed to Shuwei Wang; wangshuwei1010@foxmail.com

Received 29 December 2014; Revised 5 April 2015; Accepted 12 April 2015

Academic Editor: Jingsen Ma

Copyright © 2015 Shuwei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Along with the increasing proportion of urban public transportation trip, pedestrian flow in transportation hub areas increased. For effectively improving the emergency handling ability of related management apartments and preventing the incident of pedestrian congestion, this paper studied the method of pedestrian flow forecast in Beijing transportation hub areas. Firstly, 34 typical sidewalks in Beijing transportation hub areas were surveyed to obtain 2200 valid data. Secondly, correlation analysis was used to analyze the relationship between pedestrian flow and its influential factors. 11 significant influential factors were extracted. Thirdly, forecasting model was established with modular neural network. The surveyed pedestrian flow sample was fuzzy clustered according to the regional land use where the transportation hub existed. Then, membership function based on the distance measure was constructed. Through fuzzy discrimination, online selection for the subnetwork of the information can be achieved. Consequently, the self-adaptation of the neural network on information processing was improved. Finally, this paper tested the pedestrian flow sample of a transportation hub in Beijing. It was concluded that the accuracy of pedestrian flow forecasting model using modular neural network was higher than other neural network models. There was also improvement in the adaptability to environment.

## 1. Introduction

Recently, along with the pushing of economic development and urbanization in China, traffic congestion tended with the trend from site to line and from part to expansiveness. The related researches suggested that developing public transportation is one of the best means to solve urban traffic congestion. As a result of the rapid development of public transportation and the improving of people's environmental awareness, more and more people select public transportation as their first choice for traveling. Moreover, pedestrians in transportation hub area played an important role in affecting traffic safety and the forecast of pedestrian flow is the premise for management. Thus, research on forecasting pedestrian flow in transportation hub area is beneficial on improving the information awareness ability and emergency handling ability of related management apartments. It has an important significance on safe running of transportation hub and alleviating urban traffic congestion.

The existing researches were studied on the characteristics and forecast of pedestrian flow. Gipps and Marksjö [1] established behavior model and early route choice model of a single pedestrian on the basis of the shortest path as their moving. Hänseler et al. [2] established a cellular automaton model on pedestrians' route around railway stations based on the basic characteristics of pedestrian flow. Zhao et al. [3] surveyed on pedestrians volume in a supermarket in Harbin and fitting for a regression formula of three ordered polynomials for pedestrian flow destiny in weekdays and weekends. Li [4] established forecast models of passenger distribution and daily traffic using the passenger flow data of World Expo held in Osaka, Japan, in 1970 as main reference indicator. They also drew a curve of the passenger flow distribution. Cetiner [5] proposed an optimism dynamical neural network and on this basis built traffic flow forecast model. Marfia and Rocchetti [6] put a new forecast model into the research of the forecast of traffic congestion and examined its effectiveness with the actual results in different

TABLE 1: Correlation analysis between pedestrian flow and influential factors.

	Pearson correlation coefficient	Sig. (2-tailed)	N
Regional land usage	0.872	0.000	2200
Effective width of sidewalks	-0.296	0.000	2200
Proportion of reverse pedestrians	-0.146	0.000	2200
Type of buffer	-0.242	0.000	1257
On-street parking	0.223	0.000	2200
Isolation between nonmotor vehicles and motor vehicles	-0.072	0.002	1893
Greening	-0.134	0.000	2199
Inside motor vehicle flow	-0.168	0.000	2200
Building facilities	-0.072	0.001	2200
Distance between pedestrian and nonmotor vehicles	0.166	0.000	1893
Distance between pedestrian and motor vehicles	0.131	0.000	2200

urban roads. Xie et al. [7] proposed a hybrid temporal-spatio forecasting approach which combined temporal forecasting based on radial basis function neural network and spatio forecasting based on spatial correlation degree to obtain the passenger flow status in high-speed railway transport hub. Davidich and Köster [8] present a methodology that identifies key parameters and interdependencies that enable them to properly calibrate the model against relevant real-life data to make it capable of reproducing and predicting real-life scenarios based on the analysis of real-life data.

As analyzed above, the existing researches on pedestrian have paid much attention on pedestrians' behavior characters and passenger forecast in commercial regions and large event places. The researches on forecasting pedestrian flow were rarely. Moreover, pedestrians are highly intelligent bodies. Pedestrian flow is different from the motor vehicles which can only run on the stipulated lanes. It has the characteristics of nonlinearity and discreteness. Pedestrians can change their moving conditions according to the environment.

Aiming at solving the questions above, this paper proposed a pedestrian flow forecast model based on a modular neural network. According to the learning ability and the approximation capability of the neural network itself, the model has high precision and strong adaptive capacity to the environment. The effectiveness of the model was proved through the test on pedestrian flow survey sample in Beijing transportation hub areas.

## 2. Extraction of Significant Influence Factors

Pedestrian flow in transportation hub areas is related to the land use of the transportation hub existing area, the effective width of sidewalks, the proportion of reverse direction pedestrians, nonmotor vehicle flow, and so on. The change of climate, on-street parking, and so forth also had certain influences.

In order to reflect the situation of pedestrians in Beijing transportation hub areas and extract significant influential factors on pedestrian flow there, this paper selected 34 survey sidewalks which covered different road facilities and environmental conditions around the typical subway station in Beijing. 44 different dynamic traffic condition data were

obtained through the control survey at different buckets. The relationships between pedestrian flow and each factor were examined with correlation analysis method. Pearson correlation coefficient of the correlation analysis method is an effective method which is the most commonly used to test the relative degree between variables. It can be calculated as follows:

$$r = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right), \quad (1)$$

where  $X_i$  is the  $i$ th pedestrian flow;  $s_X$  is the variance of pedestrian flows;  $\bar{X}$  is the average of pedestrian flows;  $Y_i$  is the  $i$ th influential factor;  $s_Y$  is the variance of influential factors;  $\bar{Y}$  is the average of influential factors.

$r$  is in the range of  $-1$  and  $+1$ . If  $r > 0$ , it means that the two variables are positive correlation. If  $r < 0$ , it means that the two variables are negative correlation. The higher the absolute value of  $r$ , the stronger the relationship is. But there is no causal relationship. If  $r = 0$ , it means that the relationship between the two variables is not linear. There may be some other relations.

The correlation coefficients between pedestrian flow and each influential factor were calculated in Table 1. The Pearson correlation coefficient value of each significant influential factor at the 95% confidence interval was shown here.

As is analyzed above, 11 factors such as regional land use, effective width of sidewalks, and proportion of reverse pedestrians strongly affected pedestrian flow. Thus, the 11 factors were selected as analytical conditions to establish pedestrian flow forecast model. The model can be expressed as follows:

$$P = f(P_1, P_2, P_3, \dots, P_{11}), \quad (2)$$

where  $P$  is pedestrian flow;  $P_1$  is regional land usage;  $P_2$  is effective width of sidewalks;  $P_3$  is proportion of reverse pedestrians;  $P_4$  is type of buffer;  $P_5$  is on-street parking;  $P_6$  is isolation between nonmotor vehicles and motor vehicles;  $P_7$  is greening;  $P_8$  is inside motor vehicle flow;  $P_9$  is building facilities;  $P_{10}$  is distance between pedestrian and nonmotor vehicles;  $P_{11}$  is distance between pedestrian and motor vehicles.

The formula of pedestrian flow-time series forecast model based on chaotic time series [9] is as follows:

$$s_{t+m} = f(s_{t-1}, s_{t-2}, \dots, s_{t-q}), \quad (3)$$

where  $s_t$  is pedestrian flow at time  $t$ ;  $q$  is delay time step;  $m(> 0)$  is prediction horizon.

The essence of pedestrian flow-time series forecast formula is foreseeing the change of pedestrian flow based on the present moment and the previous information.

### 3. Integrated Method of Modular Neural Network

There are many influential factors in Beijing transportation hub area affecting pedestrian flow. The normal empirical model has certain limit in scope of application. When exceeding the scope, adoptive ability of normal empirical model is poor. However, neural network has powerful nonlinearity approximation capability and self-learning ability. It has been widely used in nonlinearity system modeling field. Thus, this paper used a neural network tool to establish pedestrian flow model.

“Urban land classification and planning construction land standard GB50137-2011” classified urban construction land into residential land, public management and service land, commercial and service facility land, industrial land, logistics storage land, traffic facility land, public facility land, and green land. The single neural network will cause high complication of the model and decrease the learning ability and generalization ability of the neural network. Modular neural network (MMN) used the theory of “divide and conquer strategy” to divide a complicating problem into several subproblems and structured local network to each subproblem. As a result, the complication of each local network was simplified and the learning ability and generalization ability of the neural network were improved. The structure of modular neural network was shown in Figure 1.

The main output of the modular neural network was

$$Y = \sum_{i=1}^K w_i y_i, \quad (4)$$

where  $Y$  is the output of the neural network;  $K$  is the classification number of sample set;  $w_i$  is weigh of the output of local network;  $y_i$  is the output of local network  $i$  ( $NET_i$ ).

According to the principle that it is similar character of the pedestrian flow in similar land use areas, this paper classified pedestrian flow based on the regional land use to achieve the purpose of task decomposition. Firstly, the feature vector of land usage was established as  $\mathbf{Z} = [RL, PMSL, CSFL, IL, LSL, TFL, PFL, GL]^T$ , where RL is residential land; PMSL is public management and service land; CSFL is commercial and service facility land; IL is industry land; LSL is logistics storage land; TFL is traffic facility land; PFL is public facility land; GL is green land. If  $\mathbf{S} = \{x_1, x_2, \dots, x_n\}$  was the observed sample set of pedestrian flow,  $n$  was the number of sample, and the set  $\mathbf{S}$  can be fuzzy

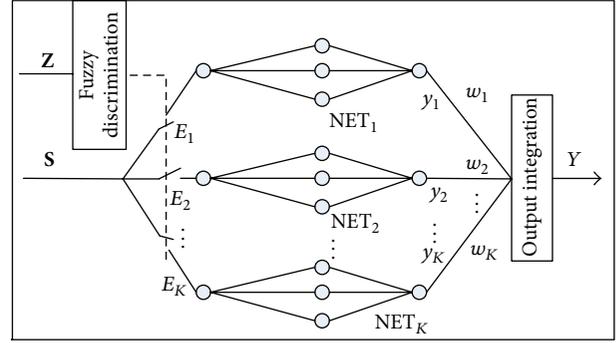


FIGURE 1: The structure diagram of a modular neural network (MNN).

clustered according to the feature vector  $\mathbf{Z}$  based on formulae (5) and (6) with the consideration of land usage. One has

$$V_i = \frac{\sum_{j=1}^N (v_{ij})^a x_j}{\sum_{j=1}^N (v_{ij})^a}. \quad (5)$$

Moreover,

$$v_{ij} = \frac{1}{\sum_{k=1}^K (\|x_j - V_i\| / \|x_j - V_k\|)^{2/(b-1)}}, \quad (6)$$

where  $\mathbf{S} = \{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_K\}$ ;  $V_i$  is clustering center of  $\mathbf{S}_i$ ;  $\mathbf{v} = (v_{ij})$  is the membership matrix of sample to a fuzzy subset. In [4],  $a = 1$ , and in [10],  $b = 2$ . Some data may simultaneously belong to sample subsets  $\mathbf{S}_i$  and  $\mathbf{S}_j$ . After that, the sample in each classification was trained in the local network that it belongs to. BP (backpropagation) learning arithmetic [11] was used to train the network.

For the new test data  $x_s \notin \mathbf{S}$ , the local network which it belongs to should be distinguished first. Distance measure can present the similarity of the two feature vectors. Thus, the distance measure of  $x_s$  and  $\mathbf{V} = [V_1, V_2, \dots, V_K]$  is chosen as a criterion to decide which local network to belong to. The calculation of distance measure was shown in formulae (7) and (8). Consider the following:

$$d_i = \frac{\|x_s - V_i\|}{\bar{d}_i}, \quad (7)$$

in which,

$$\bar{d}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \|x_j - V_i\|, \quad (8)$$

where  $d_i$  was the relative distance measure of  $x_s$  and  $NET_i$ ;  $\bar{d}_i$  is the average distance measure of  $i$ th sample subset;  $N_i$  is the number of data for  $i$ th sample subset.

The traditional modular neural network just selected the smallest local network of  $d_i$  to handle  $x_s$ . This method had a low precision in dealing with boundary samples. Consider with the fuzziness of the classification of urban land use, multiappropriate local networks joining in the handling  $x_s$

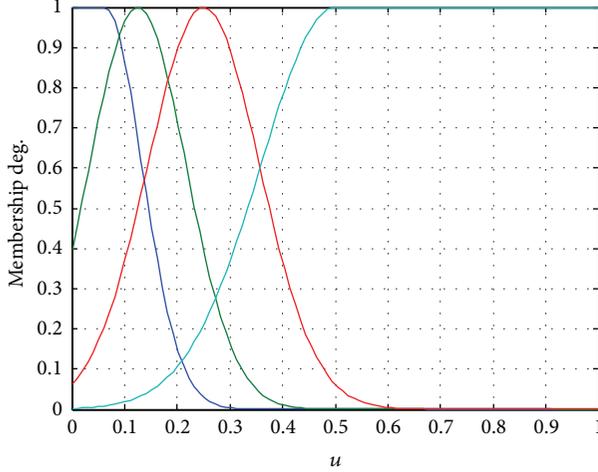


FIGURE 2: Membership degree curves of  $u_i$ .

can achieve the effect of “brainstorming” and then the error can be reduced. Thus, this paper proposed a method of selecting local network based on fuzzy decision and achieved the integrated processing to the information from multilocal networks in order to improve the accuracy of the neural network.

First, the relative distance measure of  $x_s$  among all local network normalization can be calculated using

$$u_i = \frac{d_i}{\sum_{i=1}^K d_i}. \quad (9)$$

Obviously,  $u_i \in [0, 1]$  and  $\sum u_i = 1$ .

Defining fuzzy set of distance measure  $\mathbf{A} = \{\text{very small (denoted by VS), small (denoted by S), middling (denoted by M), large (denoted by B)}\}$ . The value of  $u_i$  reflects the distance measure between  $x_s$  and  $\text{NET}_i$ . If  $u_i \in \{\text{VS}\}$ , then local networks  $\text{NET}_i$  should be used to train  $x_s$  for the distance of  $x_s$  and  $\text{NET}_i$  is short. If  $u_i \in \{\text{B}\}$ , then local network  $\text{NET}_i$  cannot be used to train  $x_s$  for the distance of  $x_s$  and  $\text{NET}_i$  is long.

The calculation of local networks of selection involves the following steps.

*Step 1.* To the input  $x_s$ , its corresponding local networks  $u_i$  values were calculated through formulas (7), (8), and (9).

*Step 2.* Calculate the membership degree of  $u_i$  to each fuzzy subset  $\mathbf{A}$  with the membership degree curves as Figure 2 shows.

*Step 3.* Make a choice which fuzzy subset  $u_i$  belongs to by the highest membership degree in the fuzzy command, and local networks which belong to the same set are chosen to integrate from VS to B. At the beginning, the local network in which  $u_i \in \{\text{VS}\}$  is selected to be trained and the others are abandoned. If the local network in which  $u_i \in \{\text{VS}\}$  does not exist, then it should be used to train the data in which  $u_i \in \{\text{S}\}$ . Otherwise, select local network in which  $u_i \in \{\text{M}\}$ .

The center of fuzzy subset to membership degree  $\{\text{VS}, \text{S}, \text{M}, \text{B}\}$  is  $0.25/K$ ,  $0.5/K$ ,  $1/K$ , and  $2/K$  and the value of membership degree is 1 when  $u_i < 0.25/K$ ,  $u_i \in \{\text{VS}\}$  or  $u_i > 2/K$ ,  $u_i \in \{\text{B}\}$ . For  $\sum u_i = 1$ , all  $u_i$  cannot meet  $u_i > 1/K$  at the same time. Meanwhile, the local network corresponding to  $u_i \in \{\text{VS}, \text{S}, \text{M}\}$  would not be the empty set. The local network corresponding to  $u_i \in \{\text{B}\}$  should not be chosen for selecting a sequence of local network from VS to B in Step 3.

The weight of local network that is selected is given by formula

$$w_{si} = \frac{1/d_{si}}{\sum_{i=1}^L 1/d_{si}}, \quad (10)$$

where  $w_{si}$  is the weight of local network that is selected;  $d_{si}$  is the distance measure of  $i$ th local network;  $L$  is the number of local networks that has been selected.

#### 4. Model Experiment Test

In order to examine the effectiveness of the model, this paper selected 34 sidewalks around transportation hub areas to test pedestrian flow. The sample was observation data (including effective width of sidewalks, proportion of reverse pedestrians, and the form of buffer) in each 15 minutes at 14:00–20:00 on April 25th to May 9th in 2013. 2400 data in total were obtained. After the quantization of the qualitative indexes such as regional land use, the form of buffer, and on-street parking, the statistical characteristics of survey data of the 12 index was obtained as showed in Tables 2 and 3. Table 2 showed the quantitative variables and Table 3 presented the qualitative ones.

Pedestrian flow was serial numbered according to the observation time order by the former to the latter. The data was classified into 8 groups based on the land use during the data preprocessing. Thus, the parameter  $K$  of modular neural network is 8. The local network used multilayered feed-forward neural network with hidden layer nodes of 4 based on the actual test. The delay time step was calculated of  $q = 6$  at the autocorrelation method.

Modular neural network model and the BP neural network passenger flow forecast model [12] both used pedestrian flow-time series method to establish a forecast model. Thus based on the methods above, the monitoring sample was trained and the accuracy of modular neural network model was examined. BP neural network pedestrian group flow forecast model mainly aimed at the region of commercial and service facility land use. 6 flow data in the former hour were used to forecast the one in the next one hour after 10 minutes.

The model was tested with 300 pedestrian flow monitoring samples in the region of commercial and service facility land use. The relative error curves of modular neural network (MNN) model and BP model were shown in Figure 3.

As is known in Figure 4, in the case of only considering commercial and service regions, the mean square error precisions for forecasting pedestrian group flow in a modular neural network model and BP neural network were both at the same level. The average relative error and the variation

TABLE 2: The statistical characteristics of the quantitative survey data of the 12 index.

Name of variables	Maximum	Minimum	Average
Pedestrian flow (person)	647	10	122
Proportion of reversing pedestrians (%)	100	0	40
Vehicle flow	389	0	79
Valid width of sidewalks (m)	10.0	1.0	2.9
Distance between pedestrians and nonmotorized vehicles (m)	4.0	0.0	1.3
Distance between pedestrians and motorized vehicles (m)	7.0	0.0	3.5
Pedestrians' satisfaction scores of the sidewalk environment (point)	10.0	1.0	5.3
Pedestrians' satisfaction scores of the inner-side facilities of sidewalks (point)	10.0	1.0	6.2

TABLE 3: The statistical characteristics of the qualitative survey data of the 12 index.

Name of variables	Description of the data
Land use characteristics	They were detailed after formula (4)
Buffer styles	No buffer, fence, greening
On-street parking	With parking, without parking
Segment between vehicles and nonmotorized vehicles	No segment, fence, greening

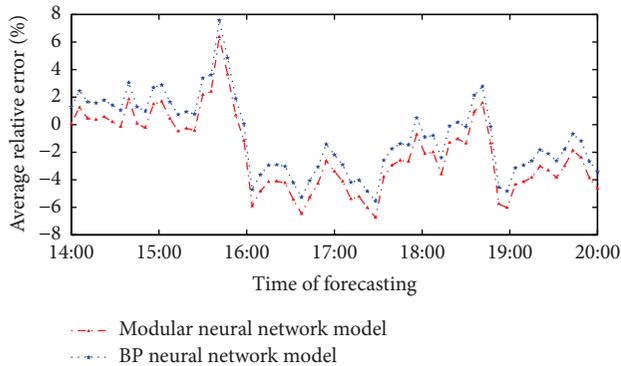


FIGURE 3: The curves of modular neural network (MNN) model and BP model.

tendency are also nearly the same. The reason was that the two models both used the BP algorithm to train the model. In the case of the same data, the error and tendency of the forecast results are also same.

Then the test on all the pedestrian flow monitoring samples in the whole monitoring cycle was held with modular neural network model and BP neural network model. The obtained curve of pedestrian group flow forecast model was shown in Figure 4.

The relative error curves of the two models were shown in Figure 5.

The mean square error and the average relative error of pedestrian group flow forecast model from modular neural network (MNN) model and BP neural network model were calculated and their results were shown in Table 4.

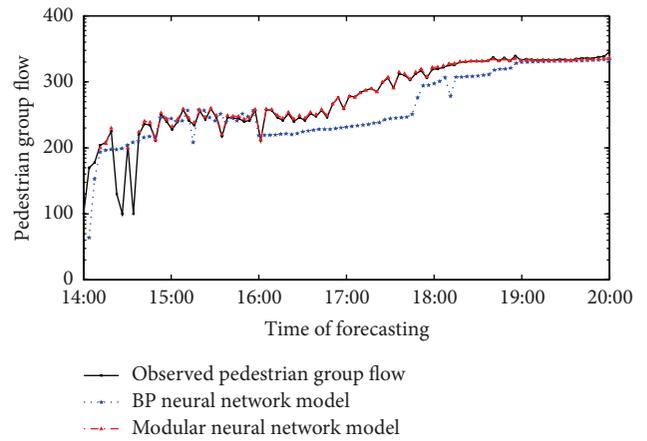


FIGURE 4: Pedestrian group flow forecast curves.

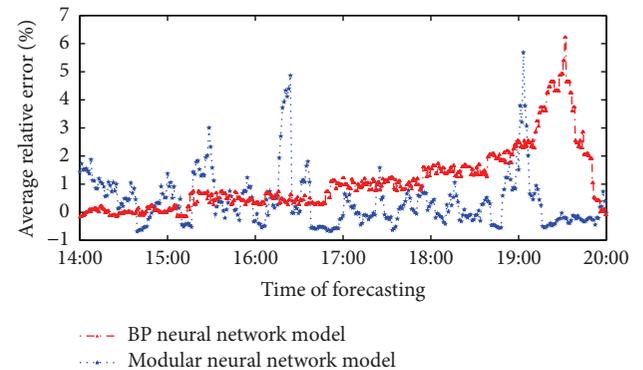


FIGURE 5: Test error curves of modular neural network (MNN) model and BP model.

From Table 4, we can obtain the following information. (1) Neural network showed the best application in the model of forecasting pedestrian flow. The accuracy of the model is high no matter which method it takes. (2) The accuracy of modular neural network forecast model is higher than the one of BP neural network pedestrian group flow forest model. In the case of total regional pedestrian flow forecast, the mean square error and the average relative error from modular neural network model are, respectively, 6.56% and 86.02% as many of the ones from the BP neural network model. (3) With the increasing of the sample, the accuracies of the two models

TABLE 4: The test error of modular neural network (MNN) model and BP model.

Model	Commercial and service region		Total region	
	Mean square error	Average relative error	Mean square error	Average relative error
Modular neural network model	$7.016 \times 10^{-2}$	3.56%	$3.662 \times 10^{-4}$	1.75%
BP neural network model	$7.531 \times 10^{-2}$	3.82%	$5.582 \times 10^{-3}$	2.03%

are both improved. However, the accuracy of the mean square error from the modular neural network model was improved by two levels and the one from BP neural network model was improve by one level. It means that modular neural model is more adapted than BP neural network model.

Based on the examination above, it can be seen that the accuracy of the forecast model of pedestrian flow in Beijing transportation hub area with the method of modular neural network (MNN) is higher than the one with BP neural network pedestrian group flow forecast model. Moreover, the traditional neural network model had a poor fitting situation if pedestrian flow exceeded the certain limit. Yet the fitting degree of the displacement value forecast model from modular neural network can stay at a high level and the model has a well-adaption.

## 5. Conclusions

- (1) Modular neural network classified pedestrian flow according to the similar character with modular neural network. It can meet the requirement of pedestrian flow forecast and deduce the complication of the neural network and then the generalization ability of the neural network is enhanced. This paper established pedestrian flow forecast model for Beijing transportation hub area based on the 11 factors which were regional land use, effective width of sidewalks, proportion of reverse pedestrians, and so on using modular neural network method. The accuracy of the model was improved to be higher than BP neural network pedestrian group flow forecast model.
- (2) Because of the learning ability and data driven features of the neural network itself, the model had a well-adoption on different environment. Along with the increasing of the samples, the adoption of modular neural network to the changing environment can be further improved with the mode of learning regularly. Then the application range of the model can be expanded.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was supported by the basic scientific research business expenses special funds from National Treasury

(no. 2015-9025). The authors are so appreciative for the participants attended this experiment.

## References

- [1] P. G. Gipps and B. Marksjö, "A micro-simulation model for pedestrian flows," *Mathematics and Computers in Simulation*, vol. 27, no. 2-3, pp. 95-105, 1985.
- [2] F. Hänseler, B. Farooq, T. Mühlematter et al., "An aggregated dynamic flow model for pedestrian movement in railway stations," in *Proceedings of the 13th Swiss Transport Research Conference*, pp. 1-14, Ascona, Switzerland, 2013.
- [3] J. N. Zhao, L. X. Wu, Z. J. Wang, and X. M. Fang, "Investigation and analysis of consumer flow rate in a large supermarket," *High Voltage Alternating Current*, vol. 34, no. 6, pp. 53-56, 2004.
- [4] K. P. Li, "Passengers flow forecasting analysis in 2010 Shanghai World Expo," *Traffic & Transportation*, no. 1, pp. 21-22, 2005.
- [5] B. G. Cetiner, "A neural network based traffic flow prediction," *Model Mathematical and Computational Applications*, vol. 15, no. 2, pp. 12-18, 2010.
- [6] G. Marfia and M. Rocchetti, "Vehicular congestion detection and short-term forecasting: a new model with results," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 2936-2948, 2011.
- [7] Z. Y. Xie, L. M. Jia, Y. Qin, and L. Wang, "A hybrid temporal-spatio forecasting approach for passenger flow status in chinese high-speed railway transport hub," *Discrete Dynamics in Nature and Society*, vol. 2013, Article ID 239039, 7 pages, 2013.
- [8] M. Davidich and G. Köster, "Predicting pedestrian flow: a methodology and a proof of concept based on real-life data," *PLoS ONE*, vol. 8, no. 12, Article ID e83355, 2013.
- [9] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica D: Nonlinear Phenomena*, vol. 73, no. 1-2, pp. 82-98, 1994.
- [10] M. G. C. A. Cimino, W. Pedrycz, B. Lazzerini, and F. Marcelloni, "Using multilayer perceptrons as receptive fields in the design of neural networks," *Neurocomputing*, vol. 72, no. 10-12, pp. 2536-2548, 2009.
- [11] Z. Q. Wang, S. F. Chen, and Z. Q. Chen, "Parallel learning approach for neural network ensemble," *Chinese Journal of Computers*, vol. 28, no. 3, pp. 402-408, 2005.
- [12] B. Sun, M. T. Li, and X. H. Yao, "Implement of crowd flow prediction based on BP neural network," *Journal of Safety Science and Technology*, vol. 6, no. 2, pp. 61-65, 2010.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

