

Research Article

Subway Sudden Passenger Flow Prediction Method Based on Two Factors: Case Study of the Dongsishitiao Station in Beijing

Chengguang Xie ¹, Xiaofeng Li ², Bingfa Chen ¹, Feng Lin ¹, Yushun Lin ¹,
and Hainan Huang ¹

¹College of Transportation and Civil Engineering, Fujian Agriculture and Forestry University, Fuzhou 350002, China

²Department of Civil and Architectural Engineering and Mechanics, The University of Arizona, Tucson 85719, AZ, USA

Correspondence should be addressed to Hainan Huang; hnan021@foxmail.com

Received 1 March 2021; Accepted 20 July 2021; Published 9 August 2021

Academic Editor: Eneko Osaba

Copyright © 2021 Chengguang Xie 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.

A sudden increase in passenger flow can primitively lead to continuous congestion of a subway network and thus have a profound impact on the subway system. To prevent the risk caused by sudden overcrowding, the prediction of passenger flow is a daily task of the rail transit management. Most current short-term passenger flow forecasts rely only on inbound passenger flow, which cannot accurately characterize the total impact of sudden passenger flow. To enhance the prediction accuracy, we propose a sudden passenger flow prediction model with two factors, the outbound and inbound passenger flows. The wavelet neural network (WNN) model was used to detect the sudden passenger flow, and subsequently, it is optimized by the genetic algorithm (GA), according to two-factor data characteristics. Sudden passenger flow events from 2014 to 2016 in the Beijing Dongsishitiao Station (DS) were used to train and verify the reliability of the prediction model. The optimized WNN results proved better than the conventional WNN, and the error of models based on two factors was significantly smaller than the models with a single-factor.

1. Introduction

A sudden passenger flow caused by large-scale activities or special holidays impact railway stations, as passengers gather in the limited station space, affecting regular operations of the station and raising the probability of risk events. On December 4, 2001, a woman on Beijing Metro Line 1 was squeezed off the platform by passengers due to the overcrowding of the platform, resulting in her death. Due to the frequent occurrence of such incidents, the Beijing Metro Department has designed protective fences on other lines to prevent passengers from being squeezed off the platform. However, there is still the risk of crowding and trampling due to large sudden passenger flows at subway stations. On November 6, 2014, at the Huixinxilu Station of Beijing Metro Line 5, a woman was crushed to death by the subway door because of overcrowding on the platform; on April 18, 2015, a trampling occurred at the Huangbeiling Station of Shenzhen Metro Line 5, which resulted in many injuries.

To prevent risks caused by the sudden crowding of passengers, the prediction of such an event has become a daily task of the rail transit management department, and then the corresponding flow control measures can be implemented based on the predicted passenger flow. In the morning and evening rush hours, multiple turning channels will be set up in the entrance square of the Beijing Subway Station to control the speed of passenger flow and ensure that the platform is not overcrowded. When there are too many passengers, some entrances will be temporarily closed to increase the difficulty of passengers entering the station. These compelling measures are all to cope with short-term passenger flow and reduce the overcrowding of subway platforms. If the occurrence time and intensity of a sudden large passenger flow can be effectively predicted, accurate limiting and diversion measures can be formulated to ensure passengers' travel safety and improve passenger travel efficiency. We propose an optimized sudden passenger flow prediction model to enhance the prediction accuracy and

aim to provide scientific forecast data for the passenger flow management of Beijing Metro.

At present, extensive research and application of short-term passenger flow prediction models are based on statistical and computational intelligence [1–3]. Three predictive models include the Bayesian network [4–6], neural networks [6–8], and a support vector machine [9–11]. However, the Bayesian network is insufficient in handling relevant input variables, the artificial neural networks (ANN) require a large amount of training data, and the support vector machine (SVM) is difficult to train and explain. In short-term astronomical passenger flow forecasting, the wavelet neural network (WNN) has shown significant advantages [12]. Jiang and Adeli [13] used dynamic wavelet analysis to build a short-term traffic flow prediction model based on neural networks and predicted actual freeway traffic flow data. Liu et al. [14] constructed a BP neural network model with traffic flow differences as input parameters and verified its feasibility, reliability, and practical value in predicting short-term traffic flow. Xia et al. [15] proposed an improved time series based on an adaptive Kalman filter (AKF) that can adjust and predict time series parameters in real time. Zhang et al. [16] established a WNN model to predict short-term passenger flow of the station, concluding that the WNN prediction accuracy was higher than the BP neural network.

Studies have shown that passengers who choose to travel by subway usually arrive and depart at the same station, and most of the sudden inbound passenger flow originates from the last outbound large passenger flow. However, most current prediction models of sudden passenger flow are based only on arrival in real time, without taking the last outbound passenger flow into consideration. Although the last outbound passenger flow was not as concentrated as the inbound passenger flow, these passengers did constitute the sudden inbound passenger flow. Therefore, this study integrated two factors, the outbound and inbound passenger flows caused by activities into the improved BP neural network model based on wavelet analysis, to enhance the learning ability and accuracy of the prediction model. Meanwhile, according to the characteristics of an orbital passenger flow time series, WNN was optimized by the genetic algorithm (GA) to solve the problems that the network computational complexity increased or the network fell into local optimum in the prediction process. Finally, data samples of sudden passenger flow events caused by activities at the Beijing Dongsishitiao Station from 2014 to 2016 were used to train and verify the reliability of the proposed model.

The remainder of this study is organized as follows. The second section introduces the optimized sudden passenger flow prediction model, the wavelet neural network, and its optimization by GA. In the third section, through the example of the Beijing Subway application, the effectiveness of the proposed model is verified. Finally, the study summarizes the research results and shortcomings.

2. Methodology

2.1. The Wavelet Neural Network Model. The wavelet neural network (WNN) model is a feedforward network based on the BP neural network and structurally indistinguishable from other feedforward networks. Compared with other neural networks, WNN is more suitable for learning the functions with local nonlinearity and rapid change.

To realize the complex nonlinear mapping problem, we selected a neural network with a hidden layer [17, 18]. Figure 1 shows the structure of the prediction model in this study. We take the sudden outbound passenger flow as the input layer, and the sudden inbound passenger flow as the output layer. The input layer x_i is the outbound passenger flow value of the i -th time node, and the output layer y_k is the sudden inbound passenger flow value of the k -th time node during the peak period. The passenger flow time serial data is processed into a symbol set data by SAX, which has been described in detail in Section 3.2.

Equation (1) describes the output of the j -th node of the hidden layer $h(j)$, where h_j is the wavelet basis function, w_{ij} is the connection function between the input layer and the hidden layer, a_j is the expansion factor, and b_j is the translation factor. The Morlet mother wavelet function is selected as h_j , as shown in equation (2).

$$h(j) = h_j \left(\frac{\sum_{i=1}^k w_{ij} x_i - b_j}{a_j} \right), \quad j = 1, 2, 3, \dots, l, \quad (1)$$

$$h_j = \cos(1.75x) e^{-x^2/2}. \quad (2)$$

Equation (3) describes the output of the k -th node of the wavelet neural network:

$$y(k) = \sum_{j=1}^l w_{jk} h(j), \quad k = 1, 2, 3, \dots, m. \quad (3)$$

2.2. Optimization by the Genetic Algorithm. The conventional WNN is a search algorithm along the gradient descent method, which makes it very sensitive to the initial values (i.e., the connection weights and thresholds). Furthermore, the search step value may lead to the risk of the result falling into a local optimum [19]. Hence, we employed the genetic algorithm (GA) to overcome the shortcomings of the WNN algorithm and improved the calculation function. Figure 2 shows the primary process, which includes the following main steps.

Step 1: Encoding. The weights and thresholds of the wavelet neural network and the scaling and translation factors of the wavelet basis function were encoded in real numbers to obtain the initial population.

Step 2: Selecting the Fitness Function. An adaptive adjustment function [20], shown in equation (4), was adopted to construct the fitness function by using the output error of the BP neural network, where ℓ is the

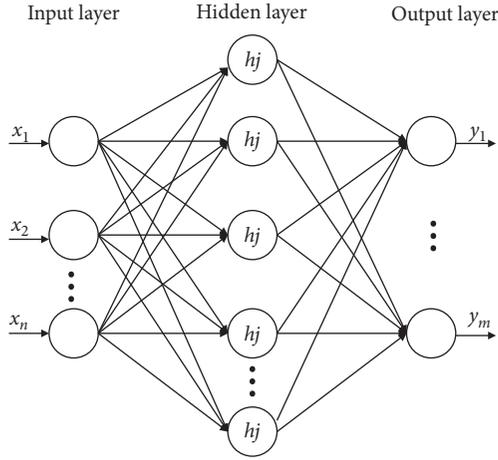


FIGURE 1: The Gplot of the wavelet neural network (WNN) model.

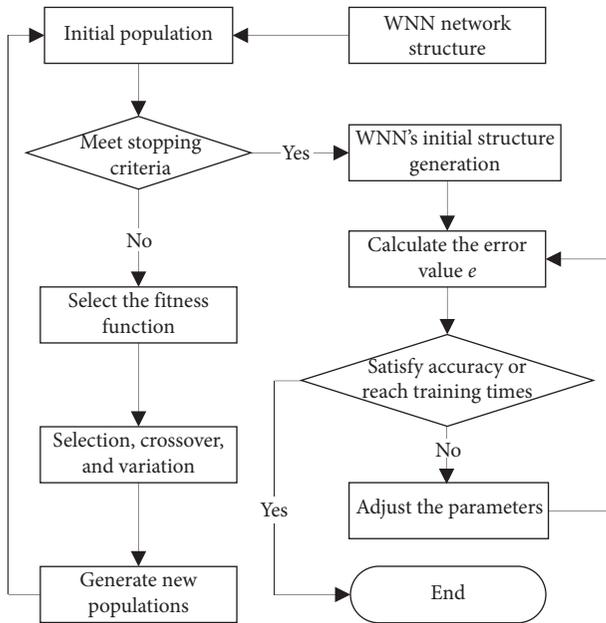


FIGURE 2: Algorithm flow of the WNN model optimized by GA.

population error, ℓ_{\max} is the maximum population error, ℓ_{\min} is the minimum population error, $\bar{\ell}$ is the average population error, and β is counted by equation (5).

$$f = \exp\left(\beta \frac{\ell_{\min}}{\ell}\right), \quad (4)$$

$$\beta = \begin{cases} 4, & 4 \leq \beta, \\ \frac{\bar{\ell}}{\ell_{\max} - \bar{\ell}}, & 1 \leq \beta < 4, \\ 1, & \beta \leq 1. \end{cases} \quad (5)$$

Step 3: *Selection, Crossover, and Variation*. According to adaptive intensity ranking, individuals with high adaptation intensity have a higher probability of

selection, and individuals with low strength have a lower selection probability. Crossover and mutation use adaptive operation, and the crossover probability p_c and the mutation probability p_m were calculated using the following equations:

$$p_c = \frac{k_1 (f_0 - f')}{f_0 - \bar{f}}, \quad (6)$$

$$p_m = \frac{k_2 (f_0 - f')}{f_0 - \bar{f}}, \quad (7)$$

where f is the fitness value, f_0 is the maximum fitness value of the population, f' is the maximum fitness value, \bar{f} is the average fitness value of the population, and k_1 and k_2 are the parameters.

Step 4: *Generating New Populations*. When a new individual was produced, it was added to the original population to produce a new population, and the cycle continued.

Step 5: *WNN's Initial Structure Generation*. If the genetic algorithm reached the number of cycles or met the preset error value, the individual was decoded, and the number of hidden layer nodes, the initial weight, the threshold, and the wavelet function parameters were obtained.

Step 6: *Calculating the Error Value e*. The sample data were input and the output error value was calculated. If the error was within the allowable range, the training ended; otherwise, it continued to step 7.

Step 7: *Adjusting the Parameters*. The error was reversed and the weight threshold, the translation factor, and the scale factor were updated.

Step 8: *The Algorithm Ending*. If the error satisfied the accuracy or reached the number of training, the training was stopped; otherwise, it continued.

3. Example Application

3.1. *Subway Data Foundation*. The Beijing Subway has a total mileage of 727 km, 24 rapid transit operating lines, and 428 stations as of 2020, as shown in the subway map in Figure 3. We chose the Dongsishitiao Station (DS) in Beijing as the research object because the DS is the nearest subway station adjacent to the Beijing Workers Stadium, where various large-scale cultural activities and sports events have been held every year. More than 30,000 people will gather in the Beijing Workers Stadium when activities are held. According to the analysis report on passenger flow characteristics of the Beijing Subway, it is relatively easy to form a sudden passenger flow at the DS.

Smart ID cards have been applied in the Beijing Subway system from 2006, and transaction records have been collected by the Automatic Fare Collection (AFC) system. As the AFC system was designed for mileage billing, subway passengers must swipe their smart ID card or scan their QR code when entering or leaving the subway station. Therefore,

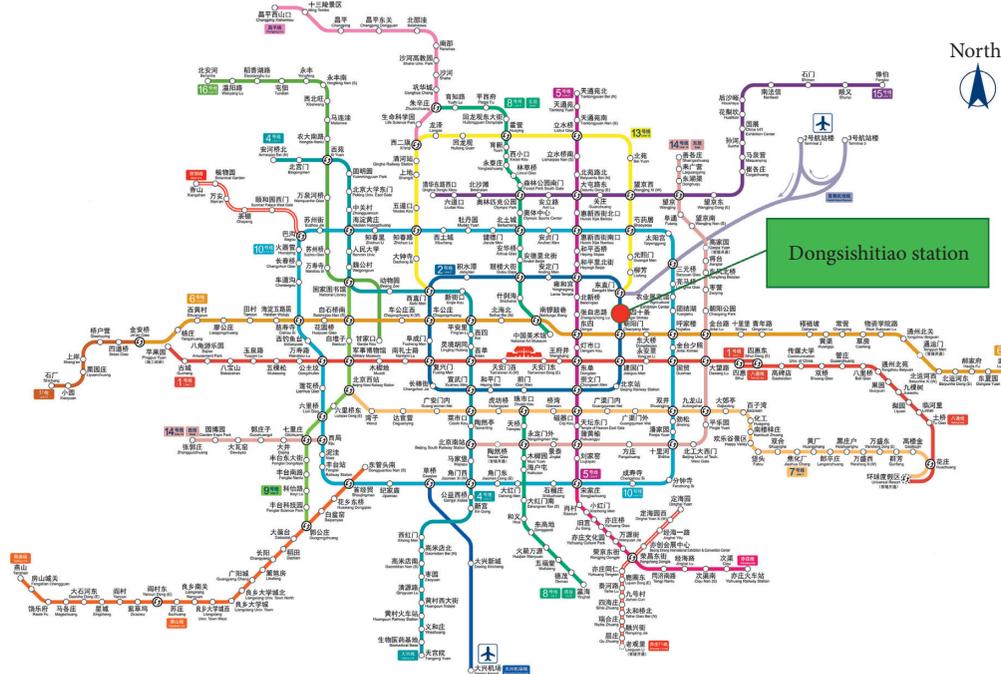


FIGURE 3: Beijing Subway map. Data source: <https://www.bjsubway.com/>.

the AFC system stores essential information, including the smart ID card number, the inbound and outbound times, the entry and exit line numbers, and the entry and exit station numbers. Furthermore, we extracted the seven core fields from the AFC transaction records of more than thirty-seven fields, as shown in Table 1. The DS's inbound and outbound data were selected from 2014 to 2016 (approximately 16 million transaction records), and the passenger flow was counted at 5 min intervals to obtain the original passenger flow time series.

3.2. Data Processing. In this study, the symbolic aggregate approximation (SAX) is employed to transfer sudden passenger flow data into a symbol set data. Considering the inbound passenger flow at a statistical interval of 5 min from a subway station as an example, a time series dimension of one-day passenger flow data is 19 (hours, 5:00 to 24:00) \times 12 (60 min/5 min). By SAX, the dimension can be converted into a symbol set with dimension 19, which significantly reduces the dimensionality of the data and improves the efficiency of the algorithm. A data processing example is shown in Figure 4 with three steps: (1) standardization in Figure 4(b), (2) dimensionality reduction in Figure 4(c), and (3) symbolization in Figure 4(d). Therefore, the time series is converted into a string $\hat{C} = [bbccbccccccffdcfcb]$.

3.3. Model Parameter Settings. It was necessary to first determine parameters and preset ranges of the algorithm process to obtain an accurate output. According to the historical passenger data analysis of DS, it was found that the duration of the sudden outbound peak was generally 3 hours (36×5 min), and the inbound peak was 2 hours

(24×5 min). Since the statistical interval was 5 min, the input layer was set to 36 nodes and the output layer to 24 nodes. According to the hidden layer node setting principle in the literature [21], the number of nodes in each hidden layer was preset from 24 to 38.

Combined with the scale and experimental analysis of the data sample of rail transit passenger flow, the selected learning rate was determined to be 0.01 to ensure that the network stability and output precision were balanced. The momentum factor was determined to be 0.9, which guaranteed the fastest drop in error.

After determining these parameters, the genetic algorithm is used to optimize the initial parameters. Figure 5 shows the result of the initial parameters optimized by GA, which indicates that (a) the optimization can finally converge; (b) the average fitness value after the 45th generation has stabilized; (c) the initial weight parameter is determined; and (d) the number of optimal hidden layer nodes is 27.

We listed the roles of several important parameters in this paper, as shown in Table 2.

3.4. Discussion. A total of 198 sudden passenger flow events caused by activities of DS were identified and used to train the proposed WNN, including 126 events in the weekend and 72 during the workdays. Meanwhile, four events in different periods were used as targets for prediction, such as July 24, 2014, March 13, 2015, October 11, 2015, and October 30, 2016. According to actual events, there were large-scale activities at the Beijing Workers Stadium on these days that generated a sizeable sudden passenger flow to DS.

Figure 6 shows the forecast results of the selected four days. The predicted time series (test) of the sudden inbound passenger flow is basically the same as the original time series

TABLE 1: Result examples of the Beijing Subway AFC system data preprocessing.

Smart card ID	Entry date and time	Entry line number	Entry station number	Exit date and time	Exit line number	Exit station number
170***	2014/5/18, 9:33	4	2	2014/5/18, 10:10	2	6
150***	2014/5/18, 19:02	10	6	2014/5/18, 19:33	8	35
154***	2014/5/18, 21:28	5	12	2014/5/18, 22:38	5	22

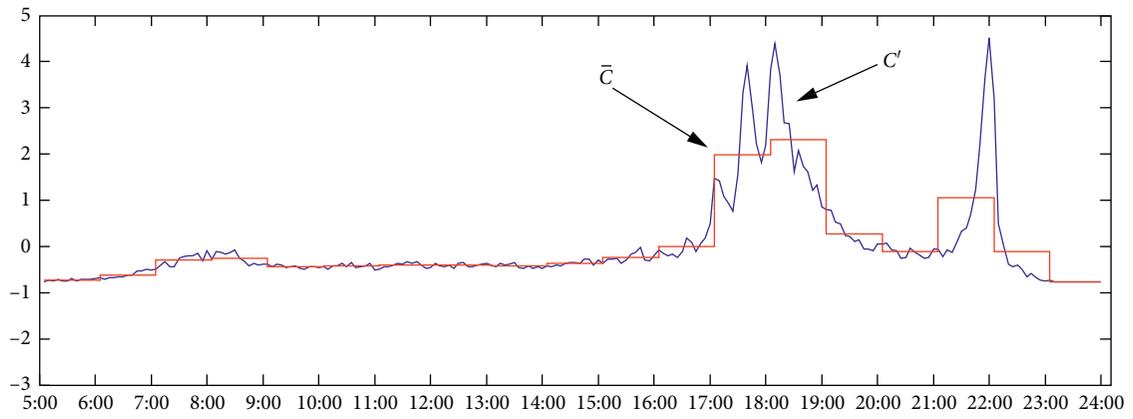
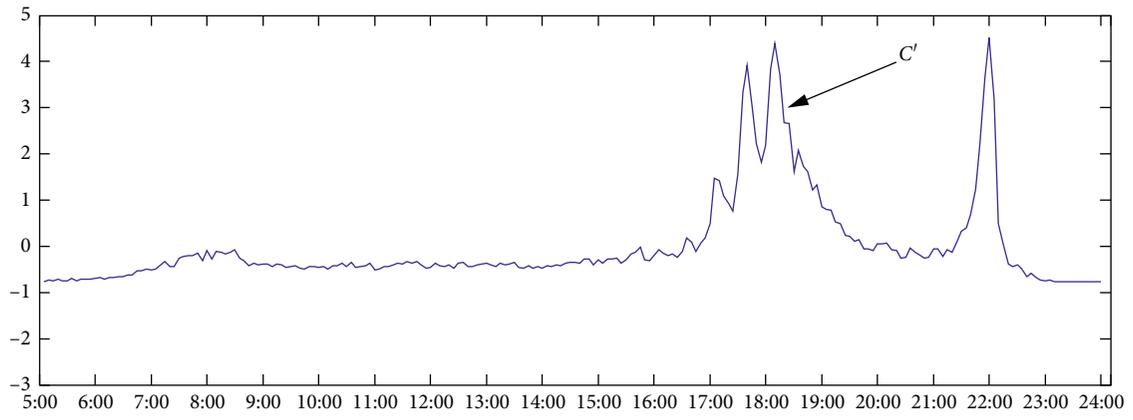
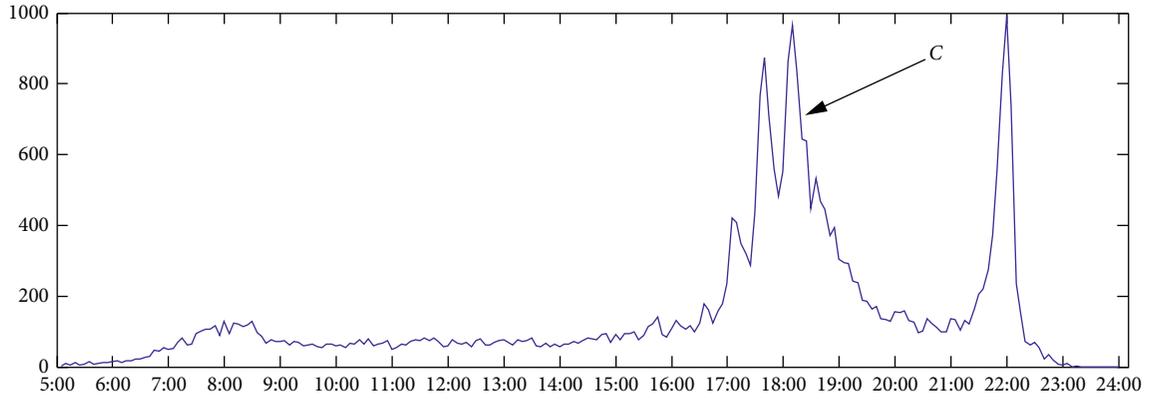
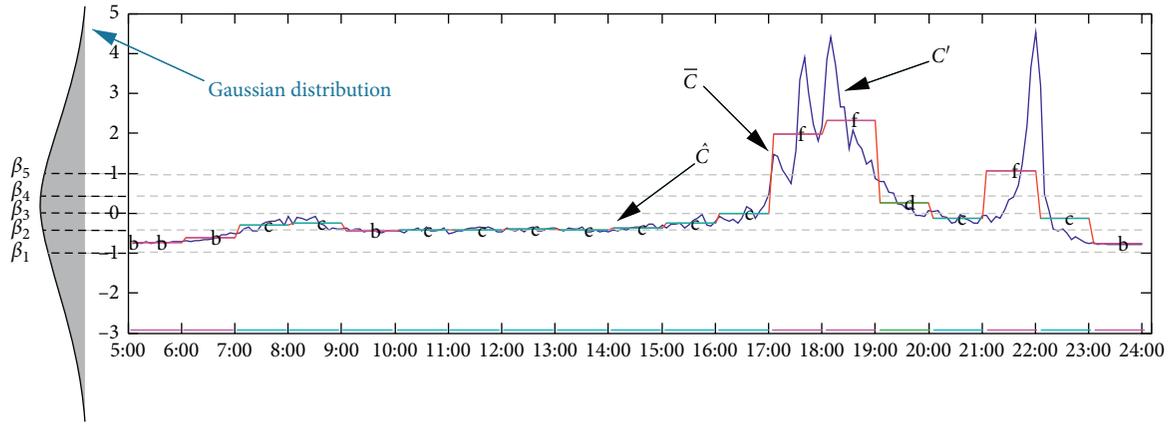


FIGURE 4: Continued.



(d)

FIGURE 4: Schematic diagram of time series symbolization. (a) Original C . (b) Standardized C' . (c) Dimensionality reduction \bar{C} . (d) Symbolized \hat{C} .

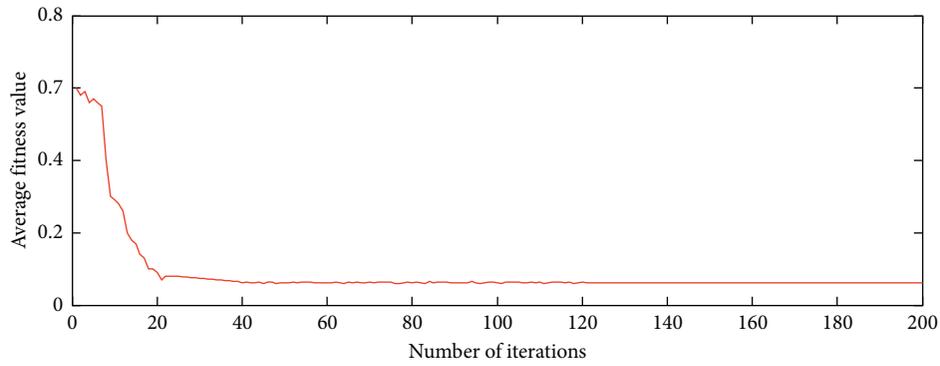
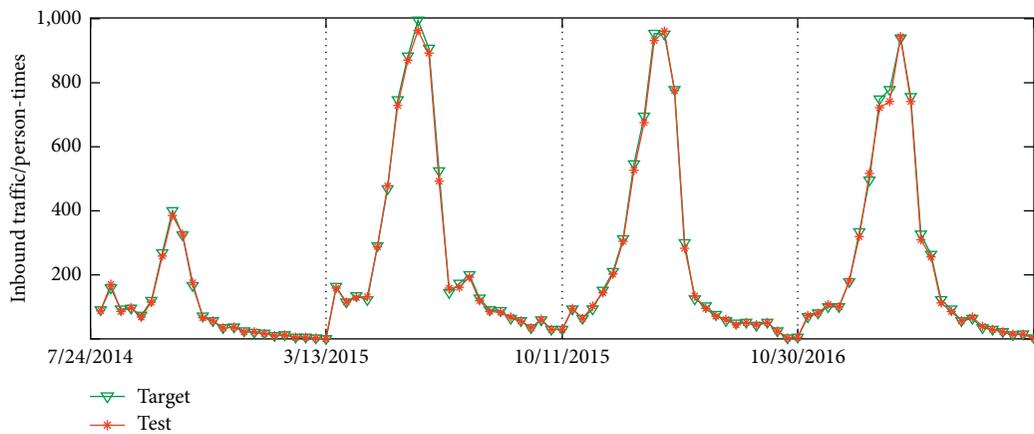


FIGURE 5: Variation of fitness of GA.

TABLE 2: Main model parameters.

Main parameters	The input layer	The output layer	Learning rate	Momentum factor	Hidden layer node
Values	36	24	0.01	0.9	27



(a)

FIGURE 6: Continued.

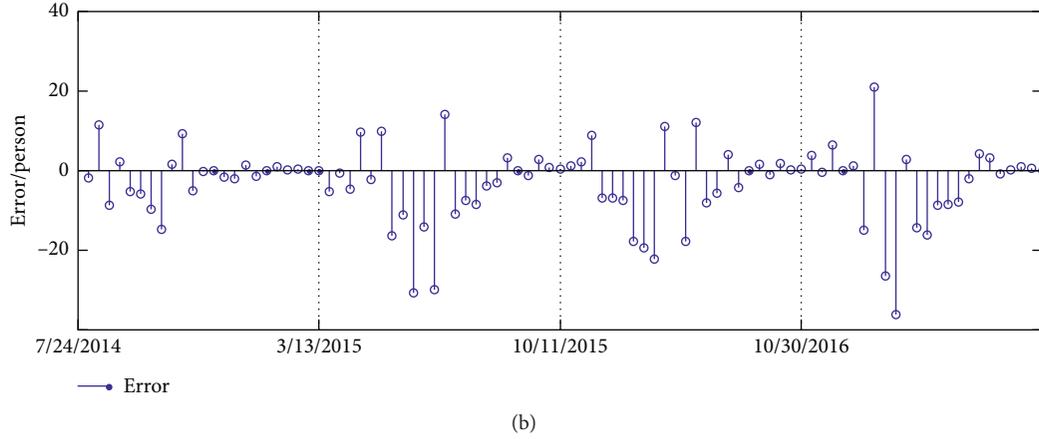


FIGURE 6: Forecast results of passenger flow caused by activities. (a) Results. (b) Error.

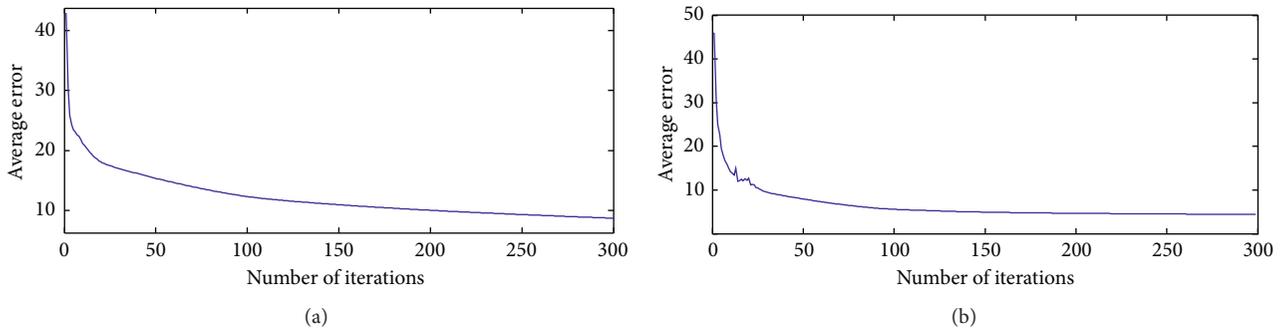


FIGURE 7: Performance curves of different forecast models. (a) WNN. (b) WNN-GA.

TABLE 3: Forecast error of using different methods, unit: person-times.

Algorithm	Factor	MSE	MAE	PME
WNN (GA)	Two	10.2	6.7	36.4
	Single	110.3	45.3	328.5
WNN	Two	48.2	23.2	84.0
	Single	128.6	60.1	462.5

(target), as shown in Figure 6(a), which indicates that the optimized prediction output value is not only numerical but also closer to the actual passenger flow in the trend.

The errors of the predicted value were calculated and are shown in Figure 6(b). Although the error of the predicted value at the peak was more significant than that of the normal passenger flow, the error value was close to 40 people, which fully met the station’s requirements for the deployment of such passenger flow restriction measures.

To better demonstrate the advantages of the two-factor predictive model, we further analyzed the prediction effects of the conventional WNN model, the optimized WNN model (by GA), the single-factor predictive model, and the two-factor predictive model. The same sample data were used to construct a conventional WNN model involving a single-inbound passenger flow, a conventional WNN model with two factors (outbound and inbound passenger flows), and an optimized WNN model (by GA) with single-inbound passenger flow.

First, the convergence speed of the conventional and optimized WNN was tested. The results shown in Figure 7 indicate that the optimized WNN model had significant improvement in convergence speed compared with the traditional WNN. In other words, the optimized WNN model (by GA) could solve the network computational complexity and local optimum problems.

Secondly, parameters such as the mean square error (MSE), the mean absolute deviation (MAE), and the peak average error (PME) of the four different models were tested. The parameters shown in Table 3 indicate that (a) the optimized WNN model by GA had the highest accuracy of the four models and (b) the models based on two factors were much better than the single-factor model, indicating that the two-factor time series played a significant role in the forecast of the subway sudden arrival passenger flow.

4. Conclusions

This study focused on the accuracy improvement of sudden arrival passenger flow prediction of the subway system. Two main improvements were implemented in this study. On the one hand, the model considered two factors, the inbound and outbound passenger flows of the subway station, which improved the accuracy of the algorithm. On the other hand, the conventional WNN model was optimized by GA to improve the efficiency of the algorithm. The example

application showed that the proposed model could effectively improve the accuracy of the prediction and had distinct advantages over the other models.

Although the results meet the requirements of the subway station staff for decision-making, it is necessary to continue to optimize the algorithm in two ways. First, as the data for the event dates of the Beijing Workers Stadium were limited and unilateral, sudden passenger flow data could be enriched with more affluent and diverse cases to better train the forecasting model. Second, the real-time prediction of sudden passenger flow should be further studied, which will be more conducive to the station management and the timeliness of shunting and limiting measures.

Data Availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: differences, similarities and some insights," *Transportation Research Part C*, vol. 19, no. 3, pp. 387–399, 2010.
- [2] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.
- [3] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, no. 1, pp. 3–19, 2014.
- [4] B. Ghosh, B. Basu, and M. O'Mahony, "Bayesian Time-Series model for Short-Term traffic flow forecasting," *Journal of Transportation Engineering*, vol. 133, no. 3, pp. 180–189, 2007.
- [5] E. Castillo, J. M. Menendez, and S. Sanchez-Cambronero, "Predicting traffic flow using Bayesian networks," *Transportation Research Part B*, vol. 42, no. 5, pp. 482–509, 2007.
- [6] W. Zheng, D.-H. Lee, and Q. Shi, "Short-Term freeway traffic flow prediction: Bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.
- [7] T. Tsai, C. Lee, and C. Wei, "Neural network based temporal feature models for short-term railway passenger demand forecasting," *Expert Systems With Applications*, vol. 36, no. 2, pp. 3728–3736, 2008.
- [8] D. Boto-Giralda, F. J. Díaz-Pernas, D. González-Ortega et al., "Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 7, pp. 530–545, 2010.
- [9] K. W. Lau and Q. H. Wu, "Local prediction of non-linear time series using support vector regression," *Pattern Recognition*, vol. 41, no. 5, pp. 1539–1547, 2007.
- [10] M. Castro-Neto, Y. Jeong, M. K. Jeong, and L. D. Han, "AADT prediction using support vector regression with data-dependent parameters," *Expert Systems With Applications*, vol. 36, no. 2, pp. 2979–2986, 2008.
- [11] N. Sapankevych and R. Sankar, "Time series prediction using support vector machines: a survey," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 24–38, 2009.
- [12] T. Li and L. Sheng, "Prediction for Short-Term traffic flow based on optimized wavelet neural network model," *International Journal of Computer Science and Information Technology*, vol. 7, no. 2, 2015.
- [13] X. Jiang and H. Adeli, "Dynamic wavelet neural network model for traffic flow forecasting," *Journal of Transportation Engineering*, vol. 131, no. 10, pp. 771–779, 2005.
- [14] Y. Liu, W. Hu, S. Xin, and L. Li, "Short-term traffic flow prediction based on genetic bp neural network," in *Proceedings of the 10th International Conference of Chinese Transportation Professionals*, Beijing, China, August 2010.
- [15] J. Xia, M. Chen, and W. Huang, "A multistep corridor travel-time prediction method using presence-type vehicle detector data," *Journal of Intelligent Transportation Systems*, vol. 15, no. 2, pp. 104–113, 2011.
- [16] X. Zhang, B. Mao, Y. Wang, J. Feng, and M. Li, "Wavelet neural network-based short-term passenger flow forecasting on urban rail transit," *Telkomnika Indonesian Journal of Electrical Engineering*, vol. 11, no. 12, pp. 7379–7385, 2013.
- [17] W. Cao, *The City Traffic Flow Prediction and Cross Analysis Based on Bp Neural Network*, Kunming University of Science and Technology, Kunming, China, 2006, in Chinese.
- [18] L. Li, B. Ran, J. Zhu, and B. Du, "Coupled application of deep learning model and quantile regression for travel time and its interval estimation using data in different dimensions," *Applied Soft Computing*, vol. 93, Article ID 106387, 2020.
- [19] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.
- [20] W. Wu, *The Study on Structural Optimization of Neural Networks Based on an Improved Genetic Algorithm*, Soochow University, Suzhou, China, 2012, in Chinese.
- [21] K. Xia, C. Li, and J. Shen, "An optimization algorithm on the number of hidden layer nodes in feed-forward neural networks," *Computer Science*, vol. 32, no. 10, pp. 143–145, 2005, In Chinese.