

CHANGXI MA, Ph.D.¹

E-mail: machangxi@mail.lzjtu.cn

LIMIN TAN, Master student¹

E-mail: 0218027@stu.lzjtu.edu.cn

XUECAI XU, Ph.D.²

(Corresponding author)

E-mail: xucai_xu@hust.edu.cn

¹ School of Traffic and Transportation

Lanzhou Jiaotong University

Lanzhou, 730070, China

² School of Civil and Hydraulic Engineering

Huazhong University of Science and Technology

Wuhan, 430074, China

Science in Traffic and Transport

Original Scientific Paper

Submitted: 18 Sep. 2019

Accepted: 25 June 2020

SHORT-TERM TRAFFIC FLOW PREDICTION BASED ON GENETIC ARTIFICIAL NEURAL NETWORK AND EXPONENTIAL SMOOTHING

ABSTRACT

In order to improve the accuracy of short-term traffic flow prediction, a combined model composed of artificial neural network optimized by using Genetic Algorithm (GA) and Exponential Smoothing (ES) has been proposed. By using the metaheuristic optimal search ability of GA, the connection weight and threshold of the feedforward neural network trained by a backpropagation algorithm are optimized to avoid the feedforward neural network falling into local optimum, and the prediction model of Genetic Artificial Neural Network (GANN) is established. An ES prediction model is presented then. In order to take the advantages of the two models, the combined model is composed of a weighted average, while the weight of the combined model is determined according to the prediction mean square error of the single model. The road traffic flow data of Xuancheng, Anhui Province with an observation interval of 5 min are used for experimental verification. Additionally, the feedforward neural network model, GANN model, ES model and combined model are compared and analysed, respectively. The results show that the prediction accuracy of the optimized feedforward neural network is much higher than that before the optimization. The prediction accuracy of the combined model is higher than that of the two single models, which verifies the feasibility and effectiveness of the combined model.

KEY WORDS

short-term traffic flow prediction; Genetic Artificial Neural Network; Exponential Smoothing; combined model;

1. INTRODUCTION

In recent years, with the development of economy and rural urbanization, the urban population has increased significantly, and the traffic demand has risen sharply. The increasing number of motor vehicles on urban roads has led to serious congestion problems in urban traffic, resulting in a large amount of energy waste and exhaust emissions. In order to alleviate urban traffic congestion more effectively, intelligent traffic control and induction has become the research focus of Intelligent Transportation System (ITS) [1-9]. Real-time, accurate and reliable short-term traffic flow prediction is the premise of intelligent traffic control and guidance. Short-term traffic flow is characterized by randomness, non-linearity and uncertainty [10]. Accurate short-term traffic flow prediction has far-reaching significance for urban transportation system. For road traffic control centre, accurate traffic flow prediction is the data support for traffic flow allocation and path optimization in the future, improving traffic induction scheme, promoting the initiative of traffic management, and avoiding traffic congestion as much as possible [11-14]. For travellers, accurate traffic flow prediction can provide them reference information so as to choose travel time and plan travel path, and improve travellers' travel convenience and comfort, as well as saving travel time. Short-term traffic flow prediction can effectively alleviate traffic congestion, which cannot only reduce

exhaust emissions and energy waste, save travel time and improve public safety, but provide theoretical support for the rapid development and application of ITS. Therefore, it has become a research hotspot in recent years.

So far, three types of models, linear models, non-linear models and combined models have been applied in short-term traffic flow prediction. Linear models mainly include historical average model [15], linear regression model [16, 17], time series model [18, 19], and Kalman filter model [20]. In historical average model, the predicted value at the current moment is the average value of historical traffic flow at the same moment at certain time interval. It is simple to calculate and only suitable for predicting the sections where traffic flow changes smoothly; The linear regression model builds the regression equation by analysing the relationship between independent variables and dependent variables, and then inputs the measured independent variable to get the predicted value. This model can be easily applied to predict the traffic flow of large-scale road network, but it is difficult for the static regression parameters to reflect the non-linear changes of traffic flow. In time series model, historical data are arranged by time series, and the predicted value is inferred by analysing the law of the change of these data over time. The time series models include autoregressive method, moving average method, autoregressive moving average method, Exponential Smoothing (ES) method, etc. Ahmed and Cook [21] firstly introduced ARIMA model for the traffic prediction area. Then, ARIMA model is widely used in highway service level prediction [22] and highway traffic flow prediction [23, 24]. Previous studies have shown that ARIMA model can reveal the static characteristics of time series, so it has high accuracy in long-term prediction, but there are limitations in short-term traffic flow prediction [25]. ES model weights time series according to smoothing coefficients and obtains linear combinations related to input data as predicted values. The modelling of time series model is simple, and the accuracy of prediction is reliable when the obtained historical data are sufficient and the data continuity are high. However, when the external factors change significantly, the accuracy of the predicted value obtained by time series analysis will decrease accordingly. The Kalman filter method describes the traffic system through the state space model composed of observation equation, state equation and Kalman filter

recursive algorithm. Okutani and Stephanedes [26] applied Kalman filter theory in dynamic short-term passenger flow prediction. Vythoulkas [27] applied Kalman filtering method to predict traffic flow in the driver information system; Zhou et al. [28] proposed a hybrid dual Kalman filter to predict short-term traffic flow. To sum up, linear model is simple in structure, weak in adaptability to various random disturbance factors, and poor in adaptability when applied to short-term prediction field, making it difficult to meet the accuracy requirements of traffic flow prediction.

Non-linear models include non-parametric regression [29], support vector machine, artificial neural network, etc. The non-parametric regression method employs the data at the next moment of the neighbouring point as the prediction value by searching for the nearest neighbours similar to the current data. Chang and Yoon [30] applied the nearest neighbour non-parametric regression method in short-term traffic flow prediction. This method requires sufficient historical data to predict more traffic flows with changing trends when there is enough data. The support vector machine method is studied according to the structural risk minimization principle. Hong [31] and Duan [32], respectively applied support vector machine models based on simulated annealing algorithm and particle swarm optimization to predict the traffic flow. Cheng et al. [33] combined the chaos theory with support vector regression model for the traffic flow prediction. Tang et al. [34] combined the data de-noising method and support vector machine for traffic flow prediction. This method requires high memory for the computer and has slower convergence speed. The artificial neural network obtains an input-output mapping relationship by training a sufficient amount of sample data, and then uses mapping to obtain a predicted value. Clark et al. [35] firstly applied artificial neural network in short-term traffic flow prediction. Subsequently, various neural networks are derived and widely used in traffic flow prediction, such as feedforward neural networks [36], dynamic Bayesian network [37], wavelet neural network [38], deep mixed neural network [39], recursive neural network [40], and long-term and short-term memory neural network [41]. Neural network can conduct self-learning, self-organization, and self-adaptation according to a large amount of data, so that it is highly suitable for random, dynamic, and non-linear traffic flow. However, there are still some

shortcomings in the neural network. For example, the learning rate of feedforward neural network trained by a traditional backpropagation algorithm is fixed, which makes the learning rate of neural network low or easy to miss the optimal solution. Feedforward neural network trained by a backpropagation algorithm relies too much on the noise of training data and is prone to over-fitting.

It can be seen from the discussion above that the single model has limitations. With the aim of fully integrating the different advantages of different models, the combination of two or more prediction models in a linear or non-linear way is called a combined model. Li et al. [42] combined ES model, ARIMA model, artificial neural network and fuzzy logic theory. Li et al. [43] combined ES method with the Markov chain; Moretti et al. [44] combined artificial neural network with linear statistical model. Hou et al. [25] combined ARIMA model with wavelet neural network. Most of the previous combined models are linear models and non-linear models. The research results show that appropriately combined models can combine the advantages of single models, overcome the limitations of single models, improve the prediction accuracy and obtain better prediction results to some extent.

According to previous research results, feedforward neural network can realize any non-linear mapping, and it has prominent advantages in short-term prediction, but it is prone to fall into local minimum value and slow learning. ES model is simple to model, but modelling parameters are generally determined by experience. In order to combine the non-linear fitting advantage of neural network with the linear fitting advantage of ES model, we propose a combined model of feedforward neural network optimized by the Genetic Algorithm (GA) and ES.

2. GA PREDICTION MODEL

The feedforward neural network has the function of realizing any complex non-linear mapping, and has strong self-learning and self-organization ability. However, the network structure cannot be easily determined, local minimum value is easy to form in the training process, and the learning speed is not high enough. To overcome the shortcomings mentioned above, this paper adopts GA to optimize the connection weights and thresholds of feedforward neural network, trains the feedforward neural network to obtain the optimal parameters, and then applies the neural network to traffic flow prediction.

2.1 Feedforward neural network

Feedforward neural network is a multi-layer perceptron network. Its topology includes the input layer, the hidden layer, and the output layer. The input information from the input layer will be transmitted to each neuron in the hidden layer, and the hidden layer will transform the information with excitation function and threshold value, and then pass the weight function to the output layer, where the information will be processed again and then output. If the output error is larger than the set value, the output error will modify the weight of the network layer by layer, until the error is smaller than the set value or the learning times reach the upper limit. The typical three-layer feedforward neural network structure is shown in *Figure 1*. Previous studies have shown that three-layer neural network can realize arbitrarily complex non-linear mapping problems [45]. For the prediction of time series, the choice of the number of input nodes and the number of hidden layer nodes is mainly determined by experience. In this paper, the number of input nodes is determined by calculating the prediction error of feedforward neural network with different input nodes. Then, the range of hidden layer nodes is determined according to *Equation 1* [46], and the optimal number of hidden layer nodes is selected according to the prediction results, where m is the number of hidden layer nodes, and n is the number of input layer nodes, and l is the number of output layer nodes.

$$m < \sqrt{n+l} + a, \quad a \in [1, 10] \tag{1}$$

In *Figure 1*, $x_i (i=1, 2, \dots, n)$ represents the input value of the first i neuron node in the input layer; $\omega_{ij} (j=1, 2, \dots, m)$ and $\omega_{jk} (k=1, 2, \dots, l)$ represent the weight between node i of the input layer and node j of the hidden layer and the weight between node j of the hidden layer and node k of the output layer,

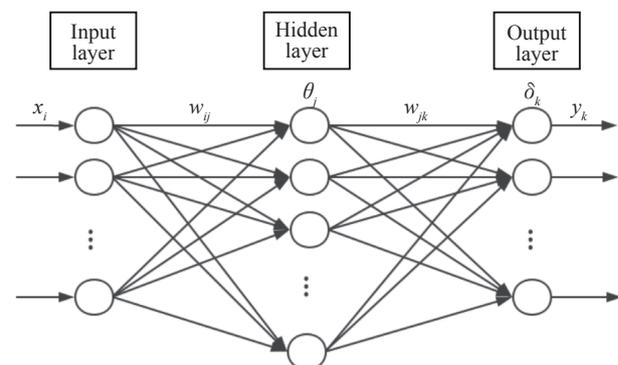


Figure 1 – Structure diagram of feedforward neural networks

respectively; θ_j denotes the threshold value of the hidden layer; δ_k denotes the output layer threshold; y_k is the output of the first k node in the output layer.

2.2 Integration artificial neural network with GA

GA is a randomized search method combining natural selection mechanism in the biological field with the Genetic theory. It combines the random exchange of metaheuristic dividable information in the population to generate new individual and the rule of survival of the fittest to achieve global optimization. GA selects the initial hypothesis solution to generate the initial population, and encodes the initial population as chromosome for selection, crossover and mutation, so as to generate new individuals. The individuals are evaluated by calculating individual fitness, and the individuals with high adaptability are selected for iteration, so as to continuously obtain better population and achieve the optimal solution [47].

When the feedforward neural network is initialized, the connection weights and thresholds of each layer are randomly valued between [-0.5,0.5]. This kind of random value method may slow down the learning speed of the neural network and get into the local optimal solution. Therefore, the optimization target of GA is set as the prediction error of the feedforward neural network. Based on the feedforward neural network determined in Section 2.1, GA is applied to optimize the selection process of weights and thresholds of the feedforward neural network, so as to obtain the optimal mapping mode and improve the prediction accuracy. The specific implementation steps are as follows:

Step 1: *Initialize parameters.* Normalize the sample data. Determine the specific topological structure of the feedforward neural network, the population size, maximum generations and other parameters of GA.
 Step 2: *Create the network.* Use the most common GA, which is binary encoding method. In the coding process, the weights of the connection between the input layer and the hidden layer, the threshold of the hidden layer, the weights of the connection between the hidden layer and the output layer and the threshold of the output layer are coded in order to form the individuals of the population. The binary string length is set as P , and the calculation formula is as follows, where m , n , and l are the node number of the hidden layer, input layer, and output layer, respectively.

$$P = nm + m + ml + l \tag{2}$$

Step 3: *Select fitness function.* This function represents the degree to which an individual approaches the optimal solution in the optimization process, and the greater the fitness, the better the individual. The goal of optimizing the feedforward neural network is to make the prediction accuracy as high as possible. Therefore, the inverse of the error between the predicted value and the expected value is selected as the fitness function, and the calculation formula is as follows:

$$f_k = \frac{1}{\sum_{k=1}^t |x_k - x'_k|} \tag{3}$$

where f_k is the fitness of an individual X_k ; t is the number of traffic flow data; x_k is the expected value, and x'_k is the predicted value.

Step 4: *Selection.* Choose the next generation of individuals according to the survival of the fittest. In other words, the next generation of individuals are selected according to their fitness in the selection. This paper selects the operator according to roulette. Roulette is also known as proportional selection, where the fitness of an individual is proportional to the chance of being selected. The calculation formula of probability p_k of individuals X_k being selected, is as follows, where d is the number of individuals.

$$P_k = \frac{f_k}{\sum_{k=1}^d f_k} \tag{4}$$

Step 5: *Crossover operation.* In accordance with the crossover probability, randomly select positions for exchange between two individuals used for reproduction. The crossover operation reflects a random exchange of information aimed at generating new gene combinations. In this paper, a simple and feasible single-point crossover operator is selected to randomly set an intersection point in an individual string for crossover. The crossover mode of the first k gene A_k and the first l gene A_l at the first j position is as follows, where the value b is randomly taken between [0,1].

$$A_{kj} = A_{kj}(1 - b) + A_{lj}b \tag{5}$$

$$A_{lj} = A_{lj}(1 - b) + A_{kj}b \tag{6}$$

Step 6: *Mutation operation.* According to the number of mutation genes produced by the mutation probability, the mutation genes are selected by a

random method. The mutation mode of the first j gene of the first i individual is as follows: where, A_{max} and A_{min} are the upper and lower bounds of the value of gene A_{ij} ; r is randomly evaluated in the interval of $[0,1]$; r' is the random number; g is the current iteration number; G_{max} is the maximum evolutionary algebra.

$$A_{ij} = \begin{cases} A_{ij} + (A_{ij} - A_{max})f(g), & r \geq 0.5 \\ A_{ij} + (A_{min} - A_{ij})f(g), & r < 0.5 \end{cases} \quad (7)$$

$$f(g) = r' \left(1 - \frac{g}{G_{max}} \right) \quad (8)$$

Step 7: *Assignment*. The optimal solution obtained by GA is considered as the initial weight and threshold of the network.

3. ES PREDICTION MODEL

ES is the most commonly used time series prediction method based on the moving average method. It roughly smooths the time series according to its historical changes, and the future trend of the time series is analysed accordingly. The two most important steps of this method are the selection of smoothing the initial value and smoothing the parameter value. However, people choose the smoothing parameter value mainly depending on experience; and once the smoothing parameter value is determined, it cannot change with the actual trend of the time series, which causes the model not to reflect the time series in real and dynamic time.

According to the number of smoothing times, three types are classified as: ES, secondary ES, and triple ES. Under normal circumstances, when the time series has no obvious trend, ES is selected; when the time series shows a linear trend, secondary smoothing is used; when the time series exhibits a non-linear trend, triple smoothing is selected. In practical applications, the trial algorithm can be selected, that is to say, according to the size of prediction error of different smoothing times, the smoothing times of corresponding time series can be determined.

Set the observation value of time series as y_1, y_2, \dots, y_t and the ES formula as

$$S_t^{(1)} = \alpha y_t + (1 - \alpha) S_{t-1}^{(1)} \quad (9)$$

$$\hat{y}_{t+1} = S_t^{(1)} \quad (10)$$

where $S_t^{(1)}$ is the ES value of the first t period; α is the smoothing coefficient, $0 < \alpha < 1$. The ES value of the first t period is the predicted value of the first $t+1$ period of the time series.

Initial ES value determination

There are three methods to determine the initial ES value ($S_0^{(1)}$): expert estimation method, the first three average methods, and the first term method.

In general, when the amount of data $n > 15$, the initial value has little effect on the predicted value, and the initial value is set as the first observation data y_1 . When $n \leq 15$, calculate the average of the first three observation data and set it as the initial value.

Selection of the smoothing parameter values

The smoothing parameter value α reflects the proportion of historical data in the ES value in different periods, that is to say, the ES method assigns weight to the historical observation value according to the influence of historical observation value on the predicted value. The closer the observation time of historical data to the prediction time, the greater the weight; otherwise, the less it weighs. Therefore, the accuracy of ES model is largely determined by the smoothing parameters. Generally, the selection of α relies on experience. When the long-term trend of time series changes slowly, α is selected between 0.1 and 0.4. When the long-term trend of time series is relatively stable, the range is 0.4-0.6. When the time series presents an obvious seasonal variation trend, α is selected between 0.6-0.9. In this paper, by calculating the Sum of Squares for Error (SSE) of the predicted value and expected value, the α which minimizes SSE is selected as a smoothing coefficient, and the calculation formula is as follows:

$$\min SSE = \sum_{k=1}^n (x_k - x'_k)^2 \quad (11)$$

4. COMBINED PREDICTION MODEL

The combined prediction model can make up for the deficiency of single model, integrate the advantages of various models, and improve the performance of prediction. According to the functional relationship between single models and the combined model, the combination method can be divided into linear combination and non-linear combination. Linear combination is a common combination prediction method with low complexity and simple calculation. The linear combination prediction model is expressed as follows;

$$Y = \omega_1 y_1 + \omega_2 y_2 \quad (12)$$

$$\omega_1 + \omega_2 = 1 \quad (13)$$

where Y is the final predicted value; ω_1 is the weight coefficient of the GANN model; ω_2 is the weight coefficient of the ES model; y_1 is the predicted value of the GANN model; and y_2 is the predicted value of the ES model.

The key factors that affect the prediction performance of linear combined model are to select the appropriate single prediction model and then determine the optimal weight coefficient of each model. The GANN model has good non-linear fitting ability and high prediction accuracy for the traffic flow with large fluctuation [48-52], while ES model has a good linear fitting ability, and it has high accuracy in predicting smooth traffic flow. Therefore, this paper establishes a combined model based on these two models (hereinafter referred to as "combined model"), which can adapt to different traffic conditions and improve the accuracy of short-term traffic flow prediction. The weight coefficient is divided into fixed weight coefficient and time-varying weight coefficient. The time-varying weight coefficient can

be dynamically adjusted to adapt to the changes of the predicted traffic flow. In this paper, the weight of the prediction results of the single model in the first T period in the combined model was determined according to the prediction root mean square error of each single model in the previous $T-1$ period, and then the weighted summation of the prediction results of all single models was used as the predicted value of the combined model in the first T period. The calculation formula of the weight coefficient $\omega_{i,T}$ of the i -th single model in period T is shown in Equation 14, where $e_{i,T}$ is the mean square error of the prediction of the first i single model in period t . Therefore, the calculation formula of predicted value Y_T during period T is shown in Equation 15. The combined model of short-term traffic flow prediction model based on GANN and ES is shown in Figure 2.

$$\omega_{i,T} = \frac{\sum_{t=1}^{T-1} e_{i,t}}{\sum_{t=1}^{T-1} \sum_{i=1}^m e_{i,t}} \tag{14}$$

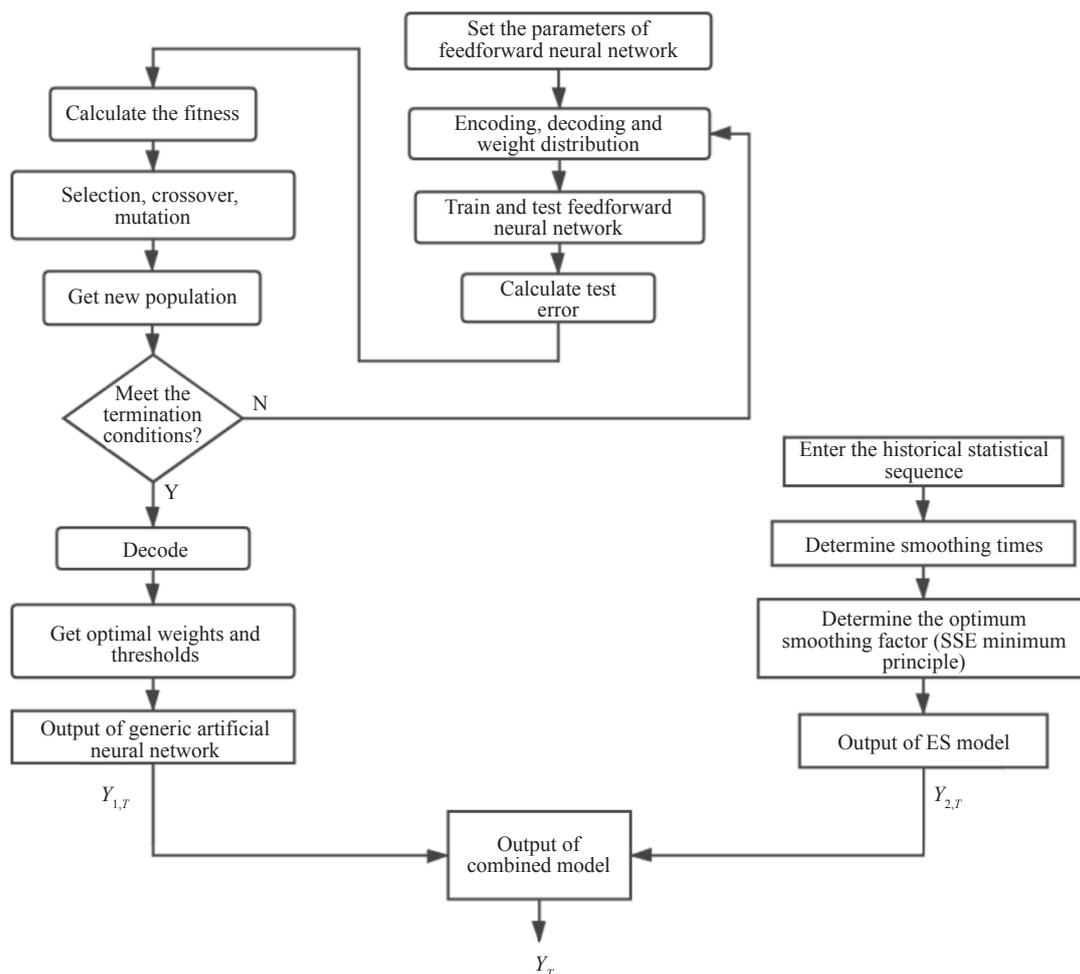


Figure 2 – Combined model

$$Y_T = \sum_{i=1}^m \left(\frac{\sum_{t=1}^{T-1} e_{i,t}}{\sum_{t=1}^{T-1} \sum_{j=1}^m e_{j,t}} \right) y_i \quad (15)$$

5. EMPIRICAL ANALYSIS

This paper analysed the measured traffic flow data of Shuangta Road, Xuancheng City, Anhui Province on December 14 and 15 (Wednesday and Thursday), 2016. Shuangta Road is an east-west main road in the city. The area around Shuangta Road is mainly residential area, so there is a large amount of commuting traffic on weekdays. Shuangta Road undertakes the important task of traffic transportation. The intersection of Shuangta Road and Zhaoting Middle Road is an important main road intersection in Xuancheng City. The traffic condition of this intersection is representative to some extent. Therefore, the traffic flow data of the west entrance section of the intersection is studied. The dataset consists of traffic flow observation data and acquisition time of the data. The sensor obtained the traffic flow observation data every five minutes, with a total of 576 data points. According to previous studies, the people's regular travel schedule leads to a certain periodicity of the traffic flow [53]. When the time period is divided into workdays, weekends and holidays, it can be found that the traffic flow presents different degrees of similarity. In case that the city layout and the scale of the road network do not change, the changes of traffic flow characteristics of a certain road are periodic. Therefore, the 576 data collected in this paper can be used as the representative of the study on traffic flow characteristics on workdays. The predicted section location is shown in *Figure 3*. The data collection station was



Figure 3 – Positions of the studied site

set at the intersection of Shuangta Road and Zhaoting Middle Road, which is marked with a triangle in *Figure 3*.

The above four models are respectively used to predict and compare the traffic flow data of the same section. MAE can reflect well the error of predicted value; RMSE can evaluate the adaptability of data and prediction model; MAPE can reflect the overall accuracy of the model. They are the commonly used indices to compare the prediction accuracy of the models, and can objectively and accurately compare the prediction effect of each model [54, 55]. Therefore, the experimental error was evaluated by MAE, MSE, RMSE and MAPE. The formula is as follows:

$$MAE = \frac{1}{n} \sum_{k=1}^n |x_k - x'_k| \quad (16)$$

$$MSE = \frac{1}{n} \sum_{k=1}^n (x_k - x'_k)^2 \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - x'_k)^2} \quad (18)$$

$$MAPE = \frac{1}{n} \sum_{k=1}^n \left| \frac{x_k - x'_k}{x_k} \right| \quad (19)$$

where n is the number of traffic flow data; x_k is the expected value; x'_k is the predicted value.

5.1 GANN prediction

Seventy percent of the 576 sample data points are selected as training data and 30% as test data. In order to limit the input and output data to a certain range and improve the speed of network training, normalization processing should be carried out before inputting data. The transformation formula is as follows:

$$x' = \frac{x_k}{x_{max}} \quad (20)$$

where x_{max} is the maximum value of the observation data; x_k is the value before the transformation; and x' is the value after the transformation. After obtaining the network output, the output should be inversely normalized. The input node number is 11 and the output node number is 1 by the method described above. Therefore, the format of the input data of GANN model is a matrix of eleven rows and one column, which contains the normalized value of eleven consecutive traffic flow data, and the output data is the traffic flow at the next time point in the interval [0,1], which needs to be reversely normalized. The number of hidden layer nodes is determined in [4,14] by *Equation 1*, and the final number of hidden layer nodes is determined from the error average of the results of ten runs. *Table 1* is the comparison of

Table 1 – Comparison of results of different number of hidden layer nodes

Number of nodes	4	6	8	10	12	14	16	18
MSE	265.2934	296.1115	304.6679	340.3700	241.3567	213.9513	260.3104	286.5470
MAE	13.0652	13.3519	14.9597	13.2223	12.2393	10.7093	11.0203	11.7892
MAPE	0.38636	0.35195	0.27387	0.23391	0.30333	0.1810	0.2034	0.2536
RMSE	16.2878	17.2079	17.4547	18.4491	15.5357	14.6271	16.1341	16.9277

Table 2 – Comparison of MSE of different transfer functions in hidden layer

Transfer function	MSE
purelin	324.0982
transig	279.4672
logsig	213.9513

the operation results of GANN model with different number of hidden layer nodes. According to Table 1, the prediction accuracy decreases firstly and then increases with the increase of the number of nodes in the hidden layer, and reaches the optimal prediction accuracy when the number of nodes is 14. In order to verify whether 14 is the best number of the hidden layer nodes, we also calculated the prediction accuracy of 16 and 18 nodes. The results show that when the number of hidden layer nodes continues to increase from 14, the prediction accuracy decreases instead of rising, presumably because the network will be overfitting when the number of hidden layer nodes is too large. Therefore, the final number of hidden layer nodes was selected as 14. The transfer function of the hidden layer was S type tangent function *transig()*. The transfer function of the output layer was selected by comparing the commonly used transfer functions of the feedforward neural network. Table 2 lists the mean MSE of the GANN model with different transfer functions in the output layer of 1,000 iterations. It can be seen from Table 2 that when the output layer adopted the S type log-

Table 3 – Comparison of prediction results of different α values

α	0.37	0.47	0.57	0.67	0.77
MSE	271.2236	259.2276	256.3527	259.0256	265.6644
MAE	11.6732	11.3835	11.3173	11.4093	11.5803
MAPE	0.1894	0.1865	0.1863	0.1879	0.1902
RMSE	16.4689	16.1005	16.0110	16.0943	16.2992

arithmic function *logsig()*, the value of MSE was the smallest. Therefore, the transfer function of the output layer was S type logarithmic function *logsig()*. *Trainlm()* function was selected as the training function, and Levenberg-Marquardt algorithm was used to train the network. The parameters of the neural network are set as follows: the number of iterations is 1,000, the training target is 0.01, and the learning rate is 0.1. When the population is initialized, the size is 40, the number of generations is 50, the crossover probability is 0.7, and the mutation probability is 0.01.

5.2 ES model prediction

The ES of time series is determined by the trial algorithm. Since there are more than 15 observation values, the first observation value is taken as $S_0^{(1)}$. By using the method of sweeping traversal search, the objective function was set as SSE of the expected value and the predicted value, and the smoothing parameter α of minimum SSE was found to be 0.57. According to experience, the smoothing parameter of time series α with a stable long-term trend was between 0.4 and 0.6, so the smoothing parameter was within a reasonable range. In order to further verify that 0.57 is the optimal value, α of 0.37, 0.47, 0.67 and 0.77 are respectively used for prediction. The comparison of prediction errors is shown in Table 3. According to Table 3, the values of MSE, MAE, MAPE and RMSE all reach the minimum when α is 0.57, and prove to be the optimal prediction result when they are 0.57.

5.3 Prediction results

The predicted results of the combined model are shown in *Figures 4* and *5*. The contrast diagram of predicted results of different models is shown in *Figure 6*. The error indices of predicted results of each model are shown in *Table 4*.

The average relative error of the feedforward neural network model is 0.1312, and the average relative error of GANN model is 0.1112, indicating that the prediction accuracy of GANN model has been significantly improved. The average relative error of the ES model is 0.1197, and the average relative error of the combined model is 0.1106, so the combined model has higher accuracy. In order to more intuitively analyse the trend of prediction results and real data, smooth de-noising is carried out for the results of the above four models and real observation data. Seen from *Figure 6*, each

model prediction results and the real observation data trend are roughly the same. In case of small fluctuations in traffic flow, the combined model is better, and the degree of fitting of GANN model is better in the traffic flow decline section, indicating that the prediction result of the combined model is not always optimal, but overall the combined model has higher prediction accuracy and more stable performance.

It can be seen from *Table 4* that the four models above can reflect better the changes of traffic flow, and the values of MAE are within 15 veh/5min, which meets the accuracy requirements of general short-term traffic flow prediction. The value of MAE of the combined model is 10.7093 veh/5min, which is lower than the other three single models. As can be seen from the RMSE error index, the error distribution of prediction results of the four models is not much different, and all of them are

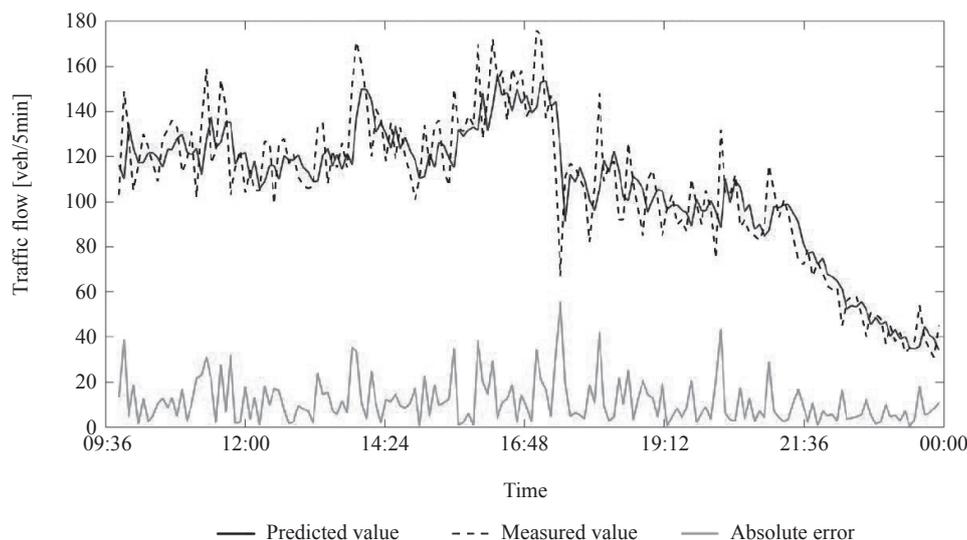


Figure 4 – The traffic flow prediction generated from the combined method

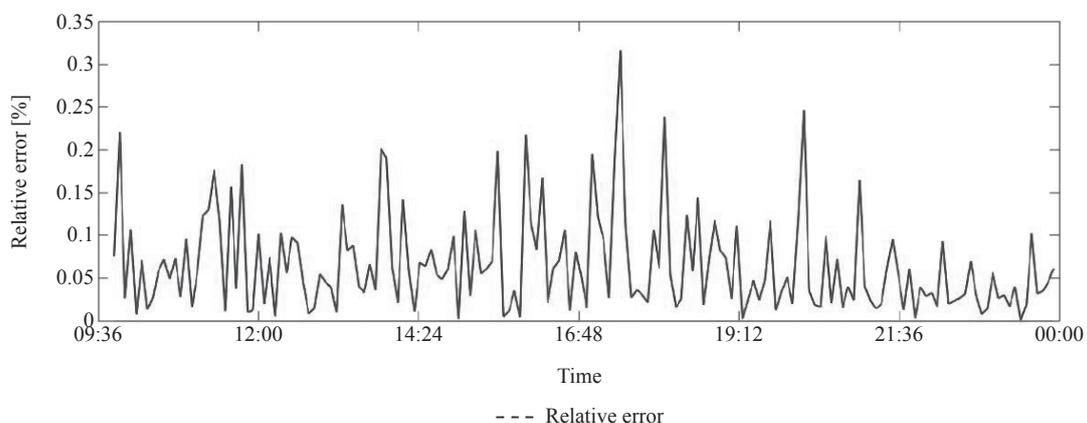


Figure 5 – Relative error of the combined method

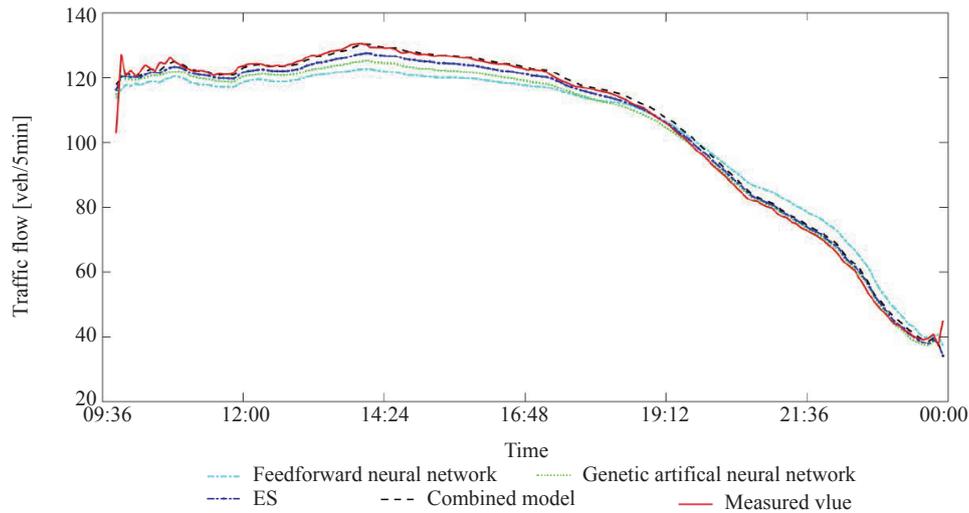


Figure 6 – Comparison of prediction results of different methods

Table 4 – The error indices of predicted results of different models

Model	MSE	MAE	MAPE	RMSE
Feedforward neural network	278.8479	12.2948	0.2356	16.6987
GANN	231.3306	11.0104	0.1971	15.2096
ES	256.3527	11.3173	0.1863	16.0110
Combined model	213.9513	10.7093	0.1810	14.6271

relatively concentrated. The prediction accuracy of the feedforward neural network optimized by GA is much higher than that of the feedforward neural network trained by a traditional backpropagation algorithm, and the value of MSE is reduced by 16.9%. Errors of the combined model are smaller than that of the single models. The value of MSE of the combined model is reduced by 7.8% compared with the GANN model and reduced by 16.8% compared with the ES model. It indicates that determining the weighting coefficients of single models according to their values of MSE can give full play to the advantages of a single model, which helps to improve the prediction accuracy. Compared with the other three single models, the combined model has better prediction results, the predicted value is closer to the real value, the prediction performance is more stable, and it has higher adaptability to different traffic flow conditions.

The prediction step was increased to predict the traffic flow in the future of 10 min and 15 min, respectively. The prediction results of the combined model are shown in Table 5. It can be seen that the prediction results of 5 min are the most accurate, and the prediction errors of 10 min and 15 min gradually increase and the accuracy gradually decreases.

In addition, as shown in Table 6, when the training sample proportion gradually decreases, MSE, MAE and MAPE gradually increase, while RMSE almost remains unchanged, indicating that the error distribution is relatively stable. In addition, the smaller the training samples are, the more the prediction accuracy of GANN model is improved than that of the feedforward neural network model. It indicates that GANN model is less restricted by the limitation of the number of training samples; in other words, it can still mine the characteristics of time series changes when the sample size decreases, which helps to maintain the accuracy of prediction. This is of significance to the small sample prediction of short-term traffic flow.

Table 5 – Comparison of the error of different step size of the combined model

Step size [min]	5	10	15
MSE	213.9513	241.2847	298.2636
MAE	10.7093	13.9872	17.2836
MAPE	0.1810	0.2173	0.2347
RMSE	14.6271	15.5333	17.2703

Table 6 – Evaluation of prediction with different training sample proportion

Training sample proportion [%]		70	60	50	40	30
MSE	Feedforward neural network	278.8479	283.8554	286.4706	287.0957	287.8540
	GANN	231.3306	233.8953	237.1431	237.9743	238.5935
	Combined model	213.9513	222.5670	228.9027	229.9231	231.8462
MAE	Feedforward neural network	12.2948	12.3736	12.6161	12.9727	13.4067
	GANN	11.0104	11.4199	11.8909	11.8929	11.9971
	Combined model	10.7093	10.7293	10.7983	10.8967	11.1776
MAPE	Feedforward neural network	0.2356	0.2438	0.2900	0.2992	0.3289
	GANN	0.1971	0.2031	0.2183	0.2233	0.2385
	Combined model	0.1810	0.1902	0.2011	0.2057	0.2168
RMSE	Feedforward neural network	16.6987	16.8480	16.9254	16.9439	16.9663
	GANN	15.2096	15.2936	15.3995	15.4264	15.4465
	Combined model	14.6271	14.9187	15.1295	15.1632	15.2265

6. CONCLUSION

In this paper, artificial neural network and ES, which are widely applied in the field of prediction, are combined to make use of their advantages in predicting non-linear data and linear data, respectively and improve the accuracy of prediction. To further improve the prediction performance of the feedforward neural network, this paper uses the strong metaheuristic optimal search ability of GA to optimize the weights and thresholds of the neural network, avoids the disadvantages brought by random assignment, and improves the prediction accuracy and convergence speed of the model.

To fully integrate the advantages of each single model, two single models are integrated by setting time-varying weight coefficient which dynamically changes when predicting error. The empirical analysis shows that the prediction accuracy of the combined model established in this paper is 7.8% and 16.8%, higher than that of the GANN model and the ES model, respectively. The fitting degree is higher, which can track the linear and non-linear change trend of traffic flow more sensitively. The model is effective and suitable for short-term traffic flow prediction.

One weakness in this research is that other factors affecting traffic flow, such as major holidays, weather conditions, etc., are not involved. In the next step, these factors will be taken into account to obtain more comprehensive traffic flow prediction data.

ACKNOWLEDGEMENTS

This work has been supported by the National Natural Science Foundation of China (Grant No. 71861023), the Program of Humanities and Social Science of Education Ministry of China (No. 18YJC630118), and the Foundation of A Hundred Youth Talents Training Program of Lanzhou Jiaotong University.

马昌喜, 博士, 教授¹

谭丽敏, 硕士生¹

徐学才, 博士, 助理教授²

¹ 兰州交通大学交通运输学院
中国兰州730070

² 华中科技大学土木与水利工程学院
中国武汉 430074

基于遗传神经网络和指数平滑算法的短时交通流预测

摘要:

为提高短时交通流预测精度, 本文采用遗传算法和指数平滑算法提出神经网络的混合算法。通过利用遗传算法的元启发式最优搜索, 采用反馈式神经网络训练的反向传播对连接权值和阈值进行优化, 以避免神经网络陷入局部最优, 因此提出遗传神经网络的预测模型和平滑指数预测模型。为充分利用两模型的优点, 混合模型由权值平均值构成, 而混合模型权值根据单个模型的预测均方差确定。实验路段采用的是安徽宣城每5分钟交通流观测值。另外, 反馈式神经网络模型, 遗传神经网络, 指数平滑模型和混合模型进行了对比, 结果显示优化的反馈式神经网络预测精度远高于优化前, 混合模型的

预测精度高于两个单一模型，验证了混合模型的可行性和有效性。

关键词：

短时交通流预测；遗传神经网络；指数平滑；混合模型

REFERENCES

- [1] Cheng R, Ge H, Wang J. An extended continuum model accounting for the driver's timid and aggressive attributions. *Physics Letters A*. 2017;381(15): 1302-1312. Available from: doi:10.1016/j.physleta.2017.02.018 [Accessed 19 June 2020].
- [2] Cheng R, Wang Y. An extended lattice hydrodynamic model considering the delayed feedback control on a curved road. *Physica A – Statistical Mechanics & Its Applications*. 2019;513: 510-517. Available from: doi:10.1016/j.physa.2018.09.014 [Accessed 19 June 2020].
- [3] Wu W, Liu R, Jin W. Designing robust schedule coordination scheme for transit networks with safety control margins. *Transportation Research Part B – Methodological*. 2016;93: 495-519. Available from: doi:10.1016/j.trb.2016.07.009 [Accessed 19 June 2020].
- [4] Ma C, Hao W, Wang A, Zhao H. Developing a coordinated signal control system for urban ring road under the vehicle-infrastructure connected environment. *IEEE Access*. 2018;6: 52471-52478. Available from: doi:10.1109/access.2018.2869890 [Accessed 19 June 2020].
- [5] Zou Y, Ash J, Park B, Lord D, Wu L. Empirical Bayes estimates of finite mixture of negative binomial regression models and its application to highway safety. *Journal of Applied Statistics*. 2018;45(9): 1652-1669. Available from: doi:10.1080/02664763.2017.1389863 [Accessed 19 June 2020].
- [6] Ma C, He R. Green wave traffic control system optimization based on adaptive genetic-artificial fish swarm algorithm. *Neural Computing and Applications*. 2019;31(7): 2073-2083. Available from: doi:10.1007/s00521-015-1931-y [Accessed 19 June 2020].
- [7] Tang J, Liang J, Zhang S, Huang H, Liu F. Inferring driving trajectories based on probabilistic model from large scale taxi GPS data. *Physica A – Statistical Mechanics and Its Applications*. 2018;506: 566-577. Available from: doi:10.1016/j.physa.2018.04.073 [Accessed 19 June 2020].
- [8] Wu W, Liu R, Jin W. Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour. *Transportation Research Part B – Methodological*. 2017;104: 175-197. Available from: doi:10.1016/j.trb.2017.06.019 [Accessed 19 June 2020].
- [9] Hao W, Ma C, Moghimi B, Fan Y. Robust optimization of signal control parameters for unsaturated intersection based on tabu search-artificial bee colony algorithm. *IEEE Access*. 2018;6: 32015-32022. Available from: doi:10.1109/access.2018.2845673 [Accessed 19 June 2020].
- [10] Li L, Qin L, Qu X, Zhang J, Wang Y, Ran B. Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm. *Knowledge Based Systems*. 2019;172: 1-14. Available from: doi:10.1016/j.knosys.2019.01.015 [Accessed 19 June 2020].
- [11] Tang J, Liang J, Han C, Li Z, Huang H. Crash injury severity analysis using a two-layer Stacking framework. *Accident Analysis & Prevention*. 2019;122: 226-238. Available from: doi:10.1016/j.aap.2018.10.016 [Accessed 19 June 2020].
- [12] Ma C, Yang D, Zhou J, Feng Z, Yuan Q. Risk riding behaviors of urban e-bikes: A literature review. *International Journal of Environmental Research and Public Health*. 2019;16(13): 2308. Available from: doi:10.3390/ijerph16132308 [Accessed 19 June 2020].
- [13] Zhang R, Ye X, Wang K, Li D, Zhu J. Development of commute mode choice model by integrating actively and passively collected travel data. *Sustainability*. 2019;11(10): 2730. Available from: doi:10.3390/su11102730 [Accessed 19 June 2020].
- [14] Wu W, Liu R, Jin W, Ma C. Stochastic bus schedule coordination considering demand assignment and re-routing of passengers. *Transportation Research Part B – Methodological*. 2019;121: 275-303. Available from: doi:10.1016/j.trb.2019.01.010 [Accessed 19 June 2020].
- [15] Stephanedes YJ, Michalopoulos PG, Plum R. Improved Estimation of Traffic Flow for Real-Time Control (Discussion and Closure). *Transportation Research Record*. 1981;(795): 28-39.
- [16] Clark SD. Traffic prediction using multivariate non-parametric regression. *Journal of Transportation Engineering – ASCE*. 2003;129(2): 161-168. Available from: doi:10.1061/(ASCE)0733-947X(2003)129:2(161) [Accessed 19 June 2020].
- [17] Turochy RE, Pierce BD. Relating Short-Term Traffic Forecasting to Current System State Using Nonparametric Regression. *Proceedings of the 7th International Conference on Intelligent Transportation Systems, 3-6 Oct 2004, Washington, USA*. New York, USA: IEEE; 2004. p. 239-244.
- [18] Hamed MM, Almasaeid HR, Said ZM. Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering – ASCE*. 1995;121(3): 249-254. Available from: doi:10.1061/(ASCE)0733-947X(1995)121:3(249) [Accessed 19 June 2020].
- [19] Kirby HR, Watson SM, Dougherty M. Should we use neural networks or statistical models for short-term motorway traffic forecasting?. *International Journal of Forecasting*. 1997;13(1): 43-50. Available from: doi:10.1016/S0169-2070(96)00699-1 [Accessed 19 June 2020].
- [20] Whittaker J, Garside S, Lindveld K. Tracking and predicting a network traffic process. *International Journal of Forecasting*. 1997;13(1): 51-61. Available from: doi:10.1016/S0169-2070(96)00700-5 [Accessed 19 June 2020].
- [21] Ahmed MS, Cook AR. Analysis of Freeway Traffic Time-series Data by Using Box-Jenkins Techniques. *Transportation Research Record*. 1979;(722).
- [22] Davis GA, Nihan NL. Using time-series designs to estimate changes in freeway level of service, despite missing data. *Transportation Research Part A: General*. 1984;18(5-6): 431-438. Available from: doi:10.1016/0191-2607(84)90018-9 [Accessed 19 June 2020].
- [23] Levin M, Tsao Y. On Forecasting Freeway Occupancies and Volumes (Abridgment). *Transportation Research Record*. 1980;(773).

- [24] Der Voort MC, Dougherty M, Watson SM. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C – Emerging Technologies*. 1996;4(5): 307-318. Available from: doi:10.1016/S0968-090X(97)82903-8 [Accessed 19 June 2020].
- [25] Hou Q, Leng J, Ma G, Liu W, Cheng Y. An adaptive hybrid model for short-term urban traffic flow prediction. *Physica A – Statistical Mechanics and Its Applications*. 2019;527: 121065. Available from: doi:10.1016/j.physa.2019.121065 [Accessed 19 June 2020].
- [26] Okutani I, Stephanedes YJ. Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B – Methodological*. 1984;18(1): 1-11. Available from: doi:10.1016/0191-2615(84)90002-X [Accessed 19 June 2020].
- [27] Vythoukcas P. Alternative Approaches to Short Term Traffic Forecasting for Use in Driver Information Systems. *Transportation and Traffic Theory*. 1993;12: 485-506.
- [28] Zhou T, Jiang D, Lin Z, Han G, Xu X, Qin J. Hybrid dual Kalman filtering model for short-term traffic flow forecasting. *IET Intelligent Transport Systems*. 2019;13(6): 1023-1032. Available from: doi:10.1049/iet-its.2018.5385 [Accessed 19 June 2020].
- [29] Ye X, Wang K, Zou Y, Lord D. A semi-nonparametric Poisson regression model for analyzing motor vehicle crash data. *PLoS One*. 2018;13(5): e0197338. Available from: doi:10.1371/journal.pone.0197338 [Accessed 19 June 2020].
- [30] Chang H, Yoon B. High-speed data-driven methodology for real-time traffic flow predictions: Practical applications of ITS. *Journal of Advanced Transportation*. 2018;2018: 1-11. Available from: doi:10.1155/2018/5728042 [Accessed 19 June 2020].
- [31] Hong W. Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. *Neurocomputing*. 2011;74(12): 2096-2107. Available from: doi:10.1016/j.neucom.2010.12.032 [Accessed 19 June 2020].
- [32] Duan M. Short-Time Prediction of Traffic Flow Based on PSO Optimized SVM. In: Huaiyin Institute of Technology and Jiangsu Key Laboratory of Advanced Manufacturing Technology. *Proceedings of 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 25-26 Jan 2018, Xiamen, China*. New York, USA: IEEE; 2018. p. 41-45.
- [33] Cheng A, Jiang X, Li Y, Zhang C, Zhu H. Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method. *Physica A – Statistical Mechanics and Its Applications*. 2017;466: 422-434. Available from: doi:10.1016/j.physa.2016.09.041 [Accessed 19 June 2020].
- [34] Tang J, Chen X, Hu Z, Zong F, Han C, Li L. Traffic flow prediction based on combination of support vector machine and data denoising schemes. *Physica A – Statistical Mechanics and Its Applications*. 2019;534: 120642. Available from: doi:10.1016/j.physa.2019.03.007 [Accessed 19 June 2020].
- [35] Clark SD, Dougherty MS, Kirby HR. The Use of Neural Networks and Time Series Models for Short Term Traffic Forecasting: A Comparative Study. *Proceedings of PTRC 21st Summer Annual Meeting, Seminar D, UMIST, Sep 1993, London, England*. London, England: PTRC; 1993.
- [36] Smith BL, Demetsky MJ. Short-Term Traffic Flow Prediction: Neural Network Approach. *Transportation Research Record*. 1994;(1453).
- [37] Roos J, Gavin G, Bonnevey S. A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data. *Transportation Research Procedia*. 2017;26: 53-61. Available from: doi:10.1016/j.trpro.2017.07.008 [Accessed 19 June 2020].
- [38] Xu L, Du X, Wang B. Short-term traffic flow prediction model of wavelet neural network based on mind evolutionary algorithm. *International Journal of Pattern Recognition and Artificial Intelligence*. 2018;32(12): 1850041. Available from: 10.1142/S0218001418500416 [Accessed 19 June 2020].
- [39] Duan Z, Yang Y, Zhang K, Ni Y, Bajgain S. Improved deep hybrid networks for urban traffic flow prediction using trajectory data. *IEEE Access*. 2018;6: 31820-31827. Available from: doi:10.1109/access.2018.2845863 [Accessed 19 June 2020].
- [40] Nejadettehad A, Mahini H, Bahrak B. Short-term demand forecasting for online car-hailing services using Recurrent Neural Networks. *arXiv: Learning*. 2019. Available from: <https://arxiv.org/abs/1901.10821v1> [Accessed 19 June 2020].
- [41] Xiao Y, Yang Y. Hybrid LSTM neural network for short-term traffic flow prediction. *Information-an International Interdisciplinary Journal*. 2019;10(3): 105. Available from: doi:10.3390/info10030105 [Accessed 19 June 2020].
- [42] Li C, Anavatti SG, Ray T. Short-term traffic flow prediction using different techniques. *Proceedings of IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society, 7-10 Nov 2011, Melbourne, Australia*. New York, USA: IEEE; 2011. p. 2423-2428.
- [43] Li J, Gao Z, Wang Z, Zhang J. Short-Term Traffic Flow Forecasting Based on Exponential Smoothing and Markov Chains. *Computer Systems & Applications*. 2013;22(12): 132-135.
- [44] Moretti F, Pizzuti S, Panziera S, Annunziato M. Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling. *Neurocomputing*. 2015;167: 3-7. Available from: doi:10.1016/j.neucom.2014.08.100 [Accessed 19 June 2020].
- [45] Li S, Liu L, Zhai M. Prediction for Short-Term Traffic Flow Based on Modified PSO Optimized BP Neural Network. *Systems Engineering – Theory & Practice*. 2012;32(9): 2045-2049.
- [46] Jiao B, Ye M. Determination of Hidden Unit Number in a BP Neural Network. *Journal of Shanghai Dianji University*. 2013;16(03): 113-116+124. Chinese
- [47] Leng Z, Gao J, Zhang B, Liu X, Ma Z. Short-Term Traffic Flow Forecasting Model of Optimized BP Neural Network Based on Genetic Algorithm. *Proceedings of the 32nd Chinese Control Conference, 26-28 July 2013, Xi'an, China*. New York, USA: IEEE; 2013. p. 8125-8129.
- [48] Niu H, Tian X, Zhou X. Demand-driven train synchronization for high-speed rail lines. *IEEE Transactions on Intelligent Transportation Systems*. 2015;16(5): 2642-2652. Available from: doi:10.1109/TITS.2015.2415513 [Accessed 19 June 2020].
- [49] Ma C, He R, Zhang W. Path optimization of taxi carpooling.

- PLoS One*. 2018;13(8): e0203221. Available from: doi:10.1371/journal.pone.0203221 [Accessed 19 June 2020].
- [50] Ma C, Hao W, Pan F, Xiang W. Road screening and distribution route multi-objective robust optimization for hazardous materials based on neural network and genetic algorithm. *PLoS One*. 2018;13(6): e0198931. Available from: doi:10.1371/journal.pone.0198931 [Accessed 19 June 2020].
- [51] Ma C, Hao W, He R, Jia X, Pan F, Fan J, Xiong R. Distribution path robust optimization of electric vehicle with multiple distribution centers. *PLoS One*. 2018;13(3): e0193789. Available from: doi:10.1371/journal.pone.0193789 [Accessed 19 June 2020].
- [52] Ma C. Network optimization design of Hazmat based on multi-objective genetic algorithm under the uncertain environment. *International Journal of Bio-Inspired Computation*. 2018;12(4): 236-244. Available from: doi:10.1504/IJBIC.2018.10017837 [Accessed 19 June 2020].
- [53] Ke J, Zheng H, Yang H, Chen X. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C – Emerging Technologies*. 2017;85: 591-608. Available from: doi:10.1016/j.trc.2017.10.016 [Accessed 19 June 2020].
- [54] Wu Y, Tan H, Qin L, Ran B, Jiang Z. A hybrid deep learning based traffic flow prediction method and its understanding. *Transportation Research Part C – Emerging Technologies*. 2018;90: 166-180. Available from: doi:10.1016/j.trc.2018.03.001 [Accessed 19 June 2020].
- [55] Li L, Qin L, Qu X, Zhang J, Wang Y, Ran B. Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm. *Knowledge Based Systems*. 2019;172: 1-14. Available from: doi:10.1016/j.knsys.2019.01.015 [Accessed 19 June 2020].