

Research Article

Prediction of Arrival Time of Pure Electric Bus Based on FA-BP Algorithm

Yuanwen Lai ¹, Hangyu Liang,¹ and Liling Huang²

¹College of Civil Engineering, Fuzhou University, Fuzhou 350116, China

²Fuzhou Urban and Rural Planning and Design Institute Company Limited, Fuzhou 350019, China

Correspondence should be addressed to Yuanwen Lai; laiyuanwen@fzu.edu.cn

Received 13 June 2022; Revised 30 May 2023; Accepted 14 June 2023; Published 10 July 2023

Academic Editor: Jing Dong-O'Brien

Copyright © 2023 Yuanwen Lai 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.

To establish a suitable pure electric bus arrival time prediction model, this paper takes pure electric bus as the research object. Based on the analysis of the influencing factors of the arrival time of the pure electric bus, the BP neural network arrival time prediction model optimized by the firefly algorithm (FA-BP prediction model) is established by selecting vehicle type, SOC value, battery age, and time as input conditions. The model is trained and tested by using bus operation data. The root mean square error of the Kalman filter model is 0.351, of the BP neural network model is 0.059, and of the FA-BP prediction model is 0.04. The results show that the model in this paper effectively improves the prediction accuracy and has good reliability and feasibility. It can provide some theoretical references for pure electric bus operation and managers and provide some basis for improving bus reliability.

1. Introduction

In recent years, the urgency of motor vehicle pollution prevention has become increasingly prominent. As a new energy vehicle, pure electric vehicles can effectively reduce air pollution, and reduce fossil energy consumption, and have been widely promoted. The wide application of pure electric vehicles in public transportation has a good demonstration and leading role and can effectively promote the development of new energy vehicles. At present, most of the existing bus travel time prediction studies are for fuel buses, natural gas buses, and hybrid buses. The operating characteristics of the pure electric bus are quite different from theirs. The existing bus arrival time prediction model and reliability model have been unable to be better applied to pure electric bus, and it is necessary to establish a suitable pure electric bus arrival time prediction model and reliability analysis model. Therefore, it is necessary to research the travel time prediction of pure electric bus.

In terms of bus arrival time prediction, to provide passengers with real-time and reliable bus arrival time, the model is generally required to have better prediction

accuracy and calculation speed. The pure electric bus is affected by weather, road conditions, and other complex factors during operation, which makes an accurate bus travel time prediction a problem. Scholars at home and abroad have invested a lot of time and energy in research to find more accurate and faster forecasting methods. Currently, the methods to predict the bus travel time include the following: historical data prediction method [1], statistical regression theory prediction model [2], time series method [3], Kalman filter model [4], artificial neural network model [5], support-vector machine model [6], probability-based prediction model [7], and particle filter-based prediction model [8]. Kumar et al. [9] proposed a generic data-driven approach for bus arrival time prediction which first learns both the spatial and temporal correlations in the historical data using supervised learning in a general nonlinear and nonstationary fashion and posed the prediction problem as a probabilistic inference problem under a nonlinear dynamical system model. Rahman et al. [10] developed a methodology to analyze the bus travel time distribution systematically based on different pseudo horizons, which takes into account the uncertainty of future bus arrival times given that early and

late buses have their respective ramifications. Huang [11], respectively, studied the influence of static factors and dynamic factors on the arrival time of the bus and established a time series prediction model based on heterogeneous information. The model is based on the recurrent neural network to study the long-term dependence of traffic and realize the prediction of bus arrival time. Serin et al. [12] handled the travel time of buses between two consecutive stops s time series, the average method, the Holt-Winters method, and the sum deep residuals method. Combining the average method, the Holt-Winters method, and the sum deep residuals method, a novel method with three-layer architecture to predict bus travel time between two stops is proposed. Achar et al. [13] utilize the historical data to learn the nonstationary (a) (linear) spatial dependencies between travel times of adjacent sections based on the above-computed order and (b) temporal dependency between successive trips as a function of the time difference between the trips and used Kalman filtering algorithm to solve the prediction model. Lai et al. [14] proposed the use of a wavelet neural network (WNN) model with an improved particle swarm optimization algorithm (IPSO) that replaces the gradient descent method. The proposed IPSO-WNN model overcomes the limitations of the gradient-based WNN, which can easily produce local optimum solutions and stop the training process and thus improving the prediction accuracy. Liu et al. [15] solve the problems of remote dependence on bus arrival and road incidents, combining the advantages of the historical data prediction method and the real-time speed data prediction method. Thus, long short-term memory and artificial neural networks on the comprehensive prediction model are proposed based on spatial-temporal feature vectors. Yang et al. [16] divided the dwell time into linear and nonlinear parts and adopted the autoregressive integrated moving average (ARIMA) model and support vector machine (SVM) to predict these two parts. Thus, the hybrid dwell time prediction method for BRT is established. Yu et al. [17] presented a hybrid model to predict bus arrival times, based on a support vector machine and Kalman filtering technique. Results show that the hybrid model generally provides better performance than artificial neural network-based methods. Yu et al. [18] proposed a random forest based on the near neighbor (RFNN) method model to predict bus travel time.

Scholars at home and abroad have conducted in-depth research on the fuel bus arrival time prediction model, but the existing research still has shortcomings. The bus arrival time prediction method does not take into account the difference in influencing factors of pure electric bus, and the prediction accuracy needs to be improved.

2. Analysis of Influencing Factors on Operation Time Reliability of Pure Electric Bus

The factors that affect the operating time of pure electric bus vehicles are complex and diverse. In addition to relatively certain factors such as driver characteristics and vehicle characteristics, random factors such as date, time period, and weather can also have an important impact on bus operating

time. Passengers are usually more concerned about the reliability and stability of bus operating time under random factors [19]. Sun [20] used a difference test to analyze the impact of date, time period, and weather on bus operation time. The results showed that random factors such as date, weather, and time period have a significant impact on bus operation time and should be considered as factors in the prediction model of bus arrival time. Pure electric buses provide operational power through battery storage of electrical energy, so battery performance determines whether pure electric buses can operate smoothly and reliably. He [21] analyzed that the battery performance of pure electric buses is mainly affected by factors such as battery capacity, battery age, and battery state of charge (SOC). Therefore, this paper selects four factors: pure electric buses vehicle type, SOC value, battery age, and time period to study their impact on operating time.

2.1. Vehicle Type. Due to different vehicle types, the vehicle quality, electric shock power, and other parameters are also different, resulting in large differences in the acceleration and deceleration performance of the bus. This paper selects bus stations with relatively close traffic volume during the off-peak period to investigate three types of pure electric buses, and the specific parameters are shown in Table 1.

Table 2 and Figure 1 show the deceleration time of the three types of pure electric bus is not very different during the deceleration process. The standard deviation of deceleration times of type A (10.7 m) and type C (7.0 m) is 1.31 s and 1.24 s, respectively, which are the maximum and minimum values of the three models. It shows that the deceleration time distribution of type C (7.0 m) is relatively uniform, while that of type A (10.7 m) is relatively dispersed. In addition, the average deceleration time of type C (7.0 m) is 6.38 s, while the average deceleration time of type B (8.0 m) and type A (10.7 m) is 6.61 s and 6.99 s, respectively. It shows that type C (7.0 m) has the best deceleration performance, followed by type B (8.0 m), and type A (10.7 m) is poor. The acceleration time characteristics of the three models of pure electric buses are similar to the deceleration time characteristics during acceleration, the standard deviation and mean of the acceleration time of type C (7.0 m) are the smallest, and the standard deviation and the average value of the acceleration time of type A (10.7 m) are the largest. It shows that the acceleration time distribution of type C (7.0 m) is relatively uniform, and the acceleration performance is the best. On the whole, type C (7.0 m) has the best acceleration and deceleration performance, followed by the B model (8.0 m), and type A (10.7 m) is poor.

Through the quantitative analysis of the relationship between acceleration and deceleration time of pure electric bus and vehicle type, it can be seen that the difference in acceleration and deceleration performance of different types of vehicles has a great difference in the impact on the bus operation time. The smaller the vehicle mass, the greater the electric shock power, and the better the acceleration and deceleration performance. Therefore, the vehicle type should be taken as an input variable of the pure electric bus arrival time prediction model.

TABLE 1: Parameter configuration of different types of pure electric bus in Fuzhou.

Bus type	A	B	C
Size (mm)	10700 × 2500 × 3070	8045 × 2350 × 2930	7000 × 2050 × 2670
Maximum speed (km/h)	69	69	100
Electric shock power (kw)	100	150	115
Total mass (kg)	18000	11400	7400

TABLE 2: Characteristic values of deceleration time of different types of pure electric bus.

Category	Whole	A	B	C	
Deceleration time	Average (s)	6.67	6.99	6.61	6.38
	Standard deviation (s)	1.29	1.31	1.25	1.24
Acceleration time	Average (s)	11.67	13.45	11.69	9.85
	Standard deviation (s)	2.68	2.53	2.12	2.04

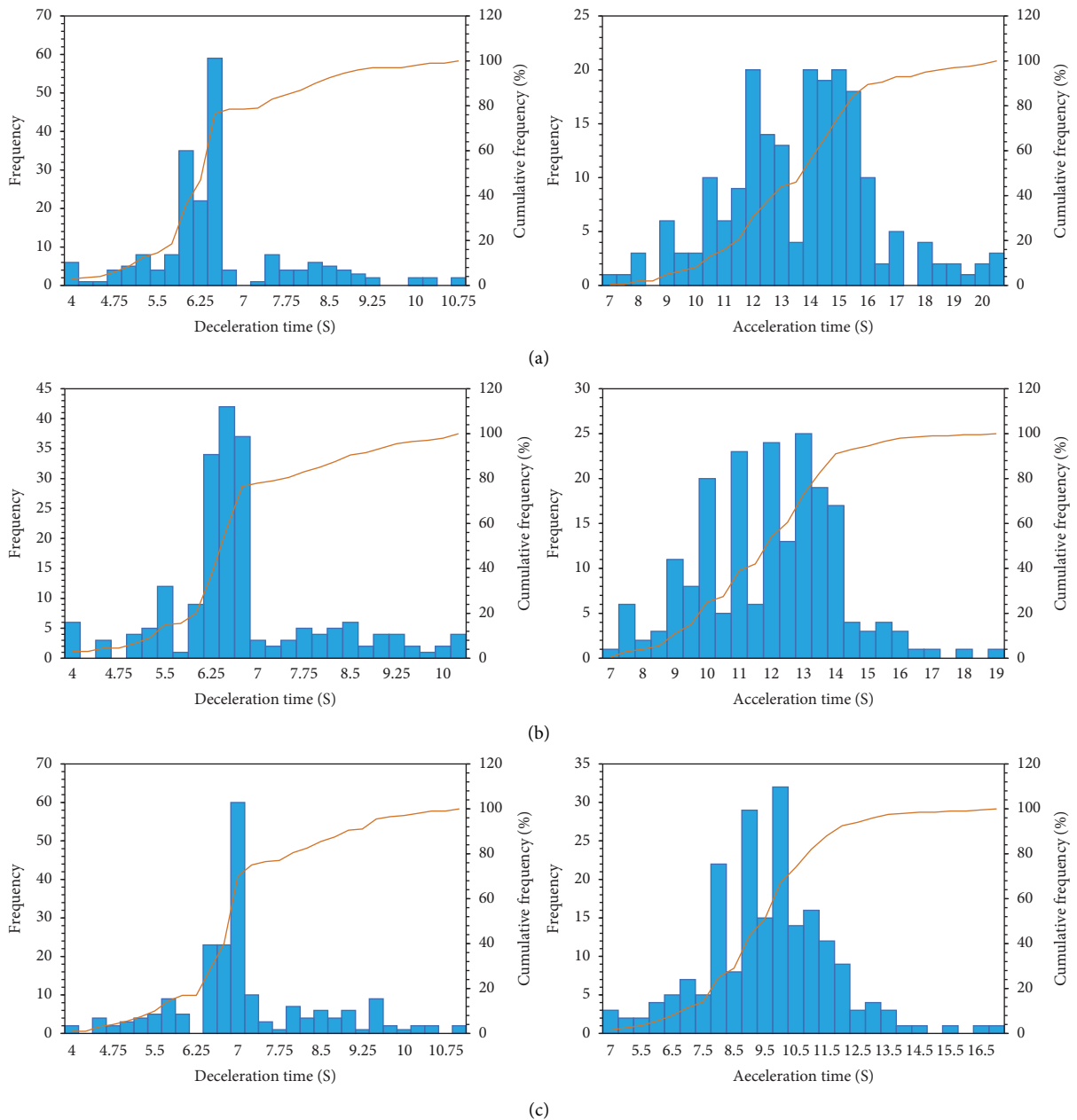


FIGURE 1: Time distribution of acceleration and deceleration of pure electric bus. (a) Bus type A. (b) Bus type B. (c) Bus type C.

2.2. SOC Value. This paper selects the operation time of the pure electric bus at Geyu Village-Youxizhou Bridgehead section of Route 321, Fuzhou City, from 14:00 to 16:00 during the off-peak period to analyze the influence of SOC value on the operation time. Figure 2 shows the change in bus operating time when the remaining power of the pure electric bus is different. In general, when the operating time of pure electric buses is between 0.4~1 SOC value, the more sufficient the power, the shorter the running time, but the difference in running time is basically between 20~30 S, and the difference is not obvious. However, when the SOC value drops below 0.4, the operating time of pure electric buses increases significantly, and the operating time of the station fluctuates greatly, and the battery is in a more unstable state. Therefore, judging the SOC value of pure electric buses will affect the arrival time of the bus.

In order to analyze whether the SOC value of pure electric buses has a significant impact on the bus running time, a random sample was selected by using ANOVA to analyze the different significance. Normality and variance homogeneity are the two prerequisites for the application of the ANOVA test, so randomly select 30 samples of the operation time of the Geyu Village-Youxizhou Bridgehead section when the SOC value is 0~0.4, 0.4~0.6, and 0.6~1.0 for testing. Due to the small amount of data in the three groups, the K-S test results of the normality tester are the mainstay. Table 3 shows the asymptotic significance of the three groups of data $> \alpha = 0.05$ so it is considered that the operation time of the three groups of pure electric buses is considered to have normality.

The homogeneity test for variance, also known as the Levene test, tests whether data from different groups come from the same population. If $> = 0.05$, the null hypothesis is accepted: the variance is homogeneous, and vice versa, the null hypothesis is rejected. Table 4 shows the test results of homogeneity of variance, P is greater than that, indicating that the above three sets of data have homogeneity of variance. Table 4 shows the test results of homogeneity of variance because $P > \alpha$, indicating that the above three groups of data have homogeneity of variance.

In Table 5, the variance of three groups of data is analyzed. Because $P = 0.001 < \alpha = 0.05$, the SOC value of pure electric buses will affect their arrival time. Therefore, it is thought that the SOC value of a pure electric bus vehicle will affect the arrival time, and the SOC value should be as an input variable of the pure electric bus arrival time prediction model.

2.3. Battery Age. Hu [22] carried out experiments on the battery under different conditions, combined with different mathematical algorithms to simulate and analyze the battery characteristics and extract the characteristic parameters.

Table 6 shows the management information of the Route 321 pure electric bus, recording the license plate number, length, battery age, and other information.

Statistics of the operation time of pure electric buses with different battery ages on the Route 321 section of Geyu Village-Youxizhou Bridgehead from 14:00 to 16:00 during the off-peak period. Figure 3 shows that, on the whole, the shorter the battery age, the shorter the operation time.

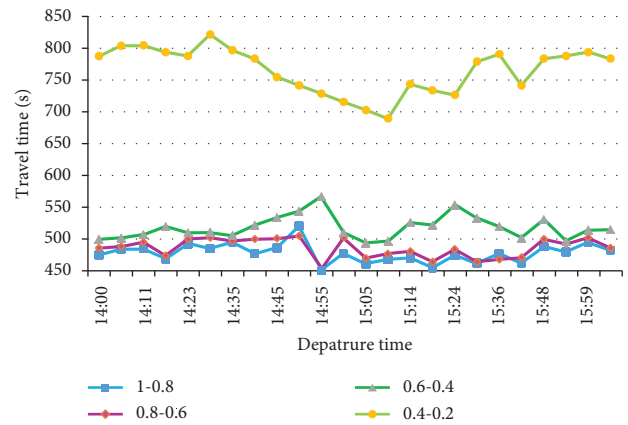


FIGURE 2: Influence of SOC value on bus travel time.

TABLE 3: Normality test of pure electric bus operation time.

SOC value	0.6~1.0	0.4~0.6	0~0.4	
N	30	30	30	
Normal parameters ^{a,b}	Mean	486.67	519.73	761.43
	Standard deviation	18.40	18.46	21.21
Kolmogorov-Smirnov	0.200	0.200	0.200	
Asymptotic significance (P value)	0.139	0.257	0.178	
Whether it obeys the normal distribution	Yes	Yes	Yes	

TABLE 4: Homogeneity test of variance.

Levene statistic	df 1	df 2	Significance (P value)
2.822	2	87	0.251

The bus arrival time corresponding to 3 years of battery age is longer than that of 0.5 years and 1.5 years of battery age, and the arrival time of buses with a battery age of 0.5 years is the shortest, so the battery age of pure electric buses is thought to affect the arrival time of the bus.

The K-S test method is used to test the normality of the above three sets of data, and the results are shown in Table 7. Because of the asymptotic significance of the three groups of data $> \alpha = 0.05$, the operation time of the three groups of pure electric buses is considered to have normality. Table 8 shows the test results of homogeneity of variances, $P > \alpha$ indicating that the three groups of operation time corresponding to batteries with different service life have homogeneity of variance.

In Table 9, the variance of the three groups of data is analyzed. Because $P = 0.001 < \alpha = 0.05$, it is believed that different battery ages of pure electric bus will affect the arrival time of pure electric buses, and the battery age should be taken as an input variable of the prediction model of the arrival time of pure electric buses.

2.4. Time Period. The impact of the time period on the travel time of pure electric buses is mainly reflected in two aspects: the density of road traffic flow and the number of passengers

TABLE 5: Analysis of variance.

	Sum of squares	df	Square of mean	F	Significance (P value)
Between groups	1350090.156	2	675045.078	1792.927	0.001
Within group	32755.900	87	376.505		
Sum	1382846.056	89			

TABLE 6: Vehicle management information of Route 321.

License plate number	Fuel type	Vehicle length	Buying time	Activation time	Battery age
Min A01166D	Pure electric	10.5	2018-01-09	2018-02-01	0.5 years
Min A01101D	Pure electric	10.5	2018-01-09	2018-02-01	0.5 years
Min AYE302	Pure electric	10.5	2016-10-31	2017-02-16	1.5 years
Min AYE310	Pure electric	10.5	2016-10-31	2017-02-16	1.5 years
Min AYD083	Pure electric	10.5	2015-11-13	2015-12-18	3 years
Min AYD718	Pure electric	10.5	2015-11-13	2015-12-18	3 years

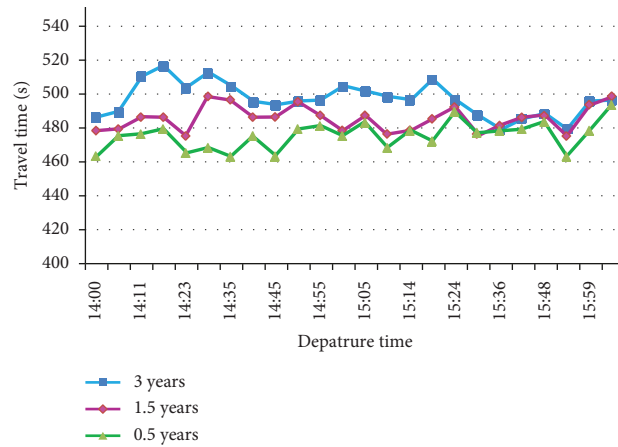


FIGURE 3: Influence of battery age of pure electric bus on bus operation time.

TABLE 7: Normality test of pure electric bus operation time.

SOC value		0.5 years	1.5 years	3 years
N		30	30	30
Normal parameters ^{a,b}	Mean	496.96	484.24	475.1
	Standard deviation	10.46	8.19	8.01
Kolmogorov-Smirnov		0.158	0.155	0.200
Asymptotic significance (P value)		0.399	0.241	0.202
Whether it obeys the normal distribution		Yes	Yes	Yes

TABLE 8: Homogeneity test of variance.

Levene statistic	df 1	df 2	Significance (P value)
2.822	2	87	0.500

boarding and alighting. When the traffic density is large, the logarithmic model proposed by Greenberg can describe the relationship between velocity and density. As shown in Figure 4, the speed of traffic flow decreases with the increase

TABLE 9: Analysis of variance.

	Sum of squares	df	Square of mean	F	Significance (P value)
Between groups	7299.267	2	3649.633	44.037	0.001
Within group	7210.333	87	82.877		
Sum	44509.600	89			

of vehicle density over time, and when the vehicle density reaches 200 veh/km, the speed of traffic flow is 0.

$$v = v_m \ln \frac{K_j}{K}, \tag{1}$$

where v_m is the maximum traffic volume speed, K is the density, and K_j is the jamming density.

The impact of the time period on the travel time of pure electric buses is mainly reflected in two aspects: road traffic flow density and the number of passengers getting on and off. During peak hours, urban residents have a high demand for commuting to and from work and school, high density of road traffic flow, and large interference of social vehicles on

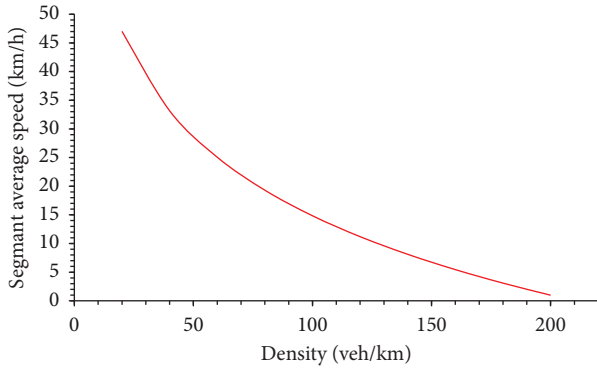


FIGURE 4: Logarithmic curve of velocity and density.

pure electric buses, resulting in low bus speeds and long travel time. At the same time, the number of bus passengers at the bus stop is large, and the time for passengers to get on and off is long, which is also one of the reasons for the long travel time of bus at peak hours. However, the road traffic flow density is relatively low in the off-peak period, the road is in a relatively smooth state, and the delay time for passengers to get on and off is also relatively short. Therefore, the travel time of pure electric buses is relatively short in the off-peak period.

Figure 5 shows the change of operation time between Geyu Village and Youxizhou Bridgehead Station on August 13, 2018, in the direction of Bus No. 321 to Yuandonglijing. Results show that the distribution of bus operation time occurs during the morning peak and evening peak between 7:00–9:00 and 17:00–19:00.

The K-S test is used to test the normality of the above two groups of data, and Table 10 shows the results. Table 10 shows the test results of the homogeneity of variance. The P values of the off-peak period and peak period are 0.057 and 0.148, respectively, which are both greater than α , indicating that the operation time of the off-peak period and peak period obeys the normal distribution.

Table 11 shows the test results of the homogeneity of variances. $P > \alpha$, indicating that the operation time of pure electric buses during peak and off-peak periods has a homogeneity of variance.

The variance of the two groups of data is analyzed in Table 12. Because $P = 0.002 < \alpha = 0.05$, it is believed that time period will affect the arrival time of the pure electric bus, and time period should be taken as an input variable of the pure electric bus arrival time prediction model.

3. Pure Electric Bus Arrival Time Prediction Model

3.1. BP Neural Networks. BP neural network is a mature and widely used error backpropagation learning algorithm in prediction, which consists of three layers: input layer, hidden layer, and output layer. There is generally no connection between neurons in each layer, and the neurons in each layer are only fully connected with neurons in adjacent layers. BP neural networks pass input signals layer by layer to the

output layer. When the obtained output value does not meet the accuracy requirements, the weight and threshold of the network are adjusted by back-propagating the error signal, and finally, the predicted value of the BP neural network gradually approaches the expected value. The BP neural network structure is shown in Figure 6.

In Figure 6, the output expression of the m neuron in the input layer of the BP neural network is

$$Y_m = f(x_m), \quad (2)$$

where f represents the conversion function of the input layer.

In Figure 6, the output expression of the p neuron in the hidden layer of the BP neural network is

$$Y_p = f\left(\sum_{m=1}^M w_{pm} Y_m - \theta_p\right), \quad (3)$$

where w_{pm} represents the connection weight between the m neuron in the input layer and the p neuron in the hidden layer; θ_p represents the hidden layer threshold.

In Figure 6, the output expression of the n neuron in the output layer of the BP neural network is

$$Y_p = f\left(\sum_{p=1}^P w_{np} Y_p - \theta_n\right), \quad (4)$$

where w_{np} represents the connection weight between the p neuron in the hidden layer and the n neuron in the output layer; θ_n represents the output layer threshold.

3.2. FA-BP Prediction Model. Considering that the travel time of the pure electric bus is nonlinear, BP neural network has a strong nonlinear mapping function and self-learning ability, which is very suitable for the prediction of bus travel time, and the prediction accuracy is higher than that of the Kalman filter, support vector machine, and other models. However, BP neural network also has the problem of a slow convergence rate and often does not continue to calculate and search after getting a nonglobal optimal solution. Therefore, the FA algorithm with global training ability and fast convergence speed is used to optimize the BP neural network model.

In addition, this paper proposes that when (1) there is no better solution within the dynamic domain of fireflies, it is stipulated that fireflies should be iterated and updated according to the guidance strategy. (2) Set the step size to a value that can be adjusted. The step size is no longer a fixed value, and its size will change with the solving process. Furthermore, the accuracy, convergence rate, and search stability of the optimal solution of the FA algorithm are improved.

3.2.1. Guided Movement Strategy. The guided movement strategy means that in the process of iteration, the optimal firefly individual is searched, and the position of the firefly is the optimal position at present. Then, the firefly will update its position according to the following formula:

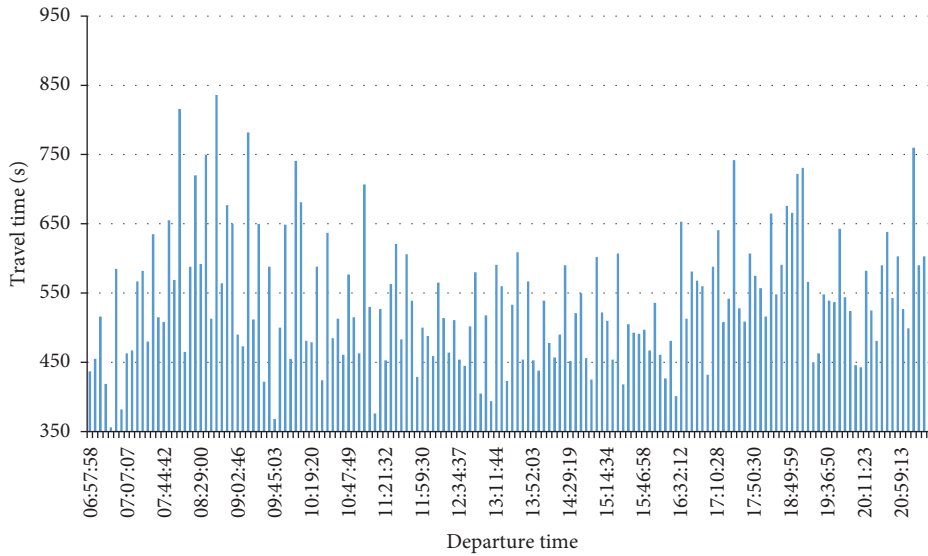


FIGURE 5: Influence of time period on the operation time of the pure electric bus.

TABLE 10: Test results of normality.

Time period	Off-peak	Peak
N	30	30
Normal parameters ^{a,b}	Mean	511.80
	Standard deviation	79.85
Kolmogorov–Smirnov Z	0.134	0.053
Asymptotic significance (P value)	0.057	0.148

TABLE 11: Homogeneity of variance test.

Levene statistic	df 1	df 2	Significance (P value)
2.466	1	58	0.122

TABLE 12: Analysis of variance.

	Sum of squares	df	Square of mean	F	Significance (P value)
Between groups	84168.817	1	94168.817	10.645	0.002
Within group	513067.767	58	8845.996		
Sum	607236.583	59			

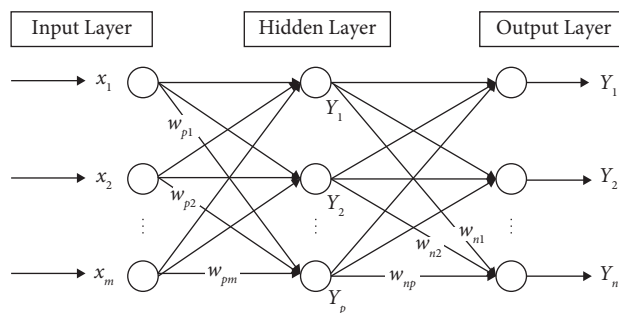


FIGURE 6: The structure of the BP neural network.

$$\begin{cases} X_i(t+1) = X_i(t) \left(\frac{X_j(t) - X_i(t)}{\|X_j(t) - X_i(t)\|} \right), \\ X_i(t+1) = X_i(t) + \text{rand}(1) * r_d^i(t), \end{cases} \quad (5)$$

where $X_i(t+1)$ represents the position of the firefly i at the $t+1$ moment; $X_i(t)$ represents the position of firefly i at time t , $s(t)$ represents the adaptive step size, $X_j(t)$ represents the position of firefly j at time t , and $r_d^i(t)$ represents the range of the dynamic decision domain of firefly i .

Firefly j has the function of guiding direction. If there is no better individual within the dynamic decision-making domain of firefly i , then firefly i will update the position in the direction of firefly k , thus greatly improving the solving rate of the firefly algorithm. If firefly i is the optimal solution in the current iteration process, it will use a smaller step size for position transformation, to achieve the purpose of searching for solutions with higher precision.

3.2.2. Adaptive Step Size Moving Strategy. Since the step size in the firefly algorithm is a constant value, a nonglobal optimal solution and repeated oscillation may occur. To solve this problem, the adaptive step size is proposed to replace the fixed step size in the algorithm. In other words, in the early iterative process of calculation, the step size should be kept large to improve the solving speed and search for the global optimal solution. In the later iterative process, the step size is kept small to solve the oscillation problem and improve the accuracy of the solution. The step size of the early and late iteration process can be determined according to equations (6) and (7), respectively.

$$s(t) = s_{\max}(p)^c, \quad (6)$$

$$c = 1 + \frac{t}{t_{\max}} - \cos\left(\frac{t}{t_{\max}}\right), \quad (7)$$

where $s(t)$ represents the adaptive step size; p is a constant.

Figure 7 shows the specific steps of the BP neural network model optimized by the firefly algorithm.

3.3. Determination of Model Input Variables. The input data to the proposed model are determined based on relevant literature (Xie [23], Lin [24], Peng and Weng [25]) and historical data on the pure electric bus's vehicle type, SOC value, battery age, and time period and bus real-time operation. The arrival time of pure electric buses is taken as the output target. The specific parameter settings are as follows.

th_n represents the operation time of the vehicle with the closest departure time from the first station of the line on the same day and the same time period in the historical week from the station i to j , which $n = 1, 2, 3$, respectively, represent the operation time data of the previous week, two weeks, and three weeks.

x_1 represents vehicle type. According to different vehicle types, the vehicle type can be expressed as $x_1 = \{1, 2, 3\}$. 1 means a bus with a mass ≥ 18000 kg, 2

means a bus with a mass of 7400 kg~18000 kg, and 3 means a bus with a mass ≤ 7400 kg.

x_2 represents SOC. According to different SOC values, SOC can be expressed as $x_2 = \{1, 2, 3\}$. 1 means that the SOC value is in the range of 0~0.4, 2 means that the SOC value is in the range of 0.4~0.6, and 3 means that the SOC value is in the range of 0.6~1.

x_3 represents the battery age. According to different battery ages, the battery age can be expressed as $x_3 = \{1, 2, 3, 4\}$. 1 means the battery age is 0~0.5 years, 2 means the battery age is 0.5~1.5 years, 3 means the battery age is 1.5~3 years, and 4 means that the battery age is more than 3 years.

x_4 represents the time period. According to the different time periods, the time period can be expressed as $x_4 = \{1, 2\}$. 1 means the peak time period of 7:00~9:00 and 17:00~19:00, and 2 means the off-peak time period of 9:00~17:00.

Input variable set $h = (th_1, th_2, th_3, x_1, x_2, x_3, x_4)^T$, namely, the seven data need to predict the arrival time of the pure electric bus.

Set up training data vector $D = (h_j, t_{ij})$. In the formula, $j = 1, 2, 3, \dots, n$, n is the number of training samples, and t_{ij} is the service time of a pure electric bus from the station i to j .

3.4. Input Data Processing. This paper uses the normalized function as formula (2):

$$X^* = \frac{x_i - 1/K \sum_{i=1}^K x_i}{x_{\max} - x_{\min}}, \quad (8)$$

where X^* represents the value after normalization; K represents the number of groups of training data; x_{\max} represents the maximum value of training data; x_{\min} represents the minimum value of training data.

3.5. Selection of Activation Function. Figure 8 shows that the value ranges of the bipolar S-function and the S-function are different. The minimum value of the S-shaped function is 0, the maximum value is 1, and the value range is 0~1. The minimum value of the bipolar S-function is -1, the maximum value is 1, and the value range is -1~1. Because the bus travel time cannot be negative, this paper chooses the S-shape function as the activation function of the BP neural network.

S-type function:

$$f(x) = \frac{1}{1 + e^{-ax}}. \quad (9)$$

3.6. Determination of the Number of Model Nodes. The number of nodes in the input layer is the number of influencing factors, so the number of input layers is 7. In this model, the operation time of the pure electric bus is the only output value of the output layer, so the number of nodes in the output layer is 1. The number of nodes in the hidden layer is determined as shown in the following formula:

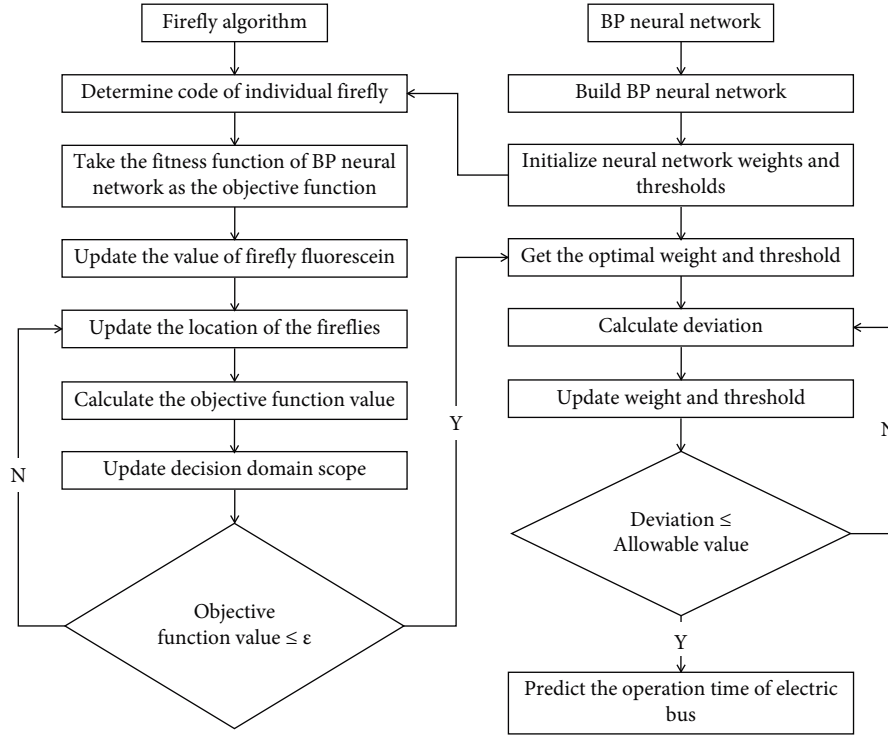


FIGURE 7: Steps of FA-BP prediction model.

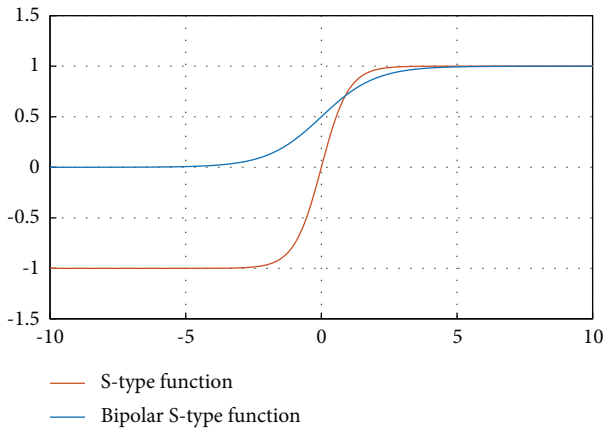


FIGURE 8: S-type function and bipolar S-type function.

$$P < \sqrt{M + N} + \lambda, \tag{10}$$

where P represents the number of nodes in the hidden layer, M represents the number of nodes in the input layer, N represents the number of nodes in the output layer, and λ is a constant between 0 and 10.

Through repeated training, when the number of hidden layer nodes is 9, the prediction accuracy of the model is highest.

4. Example Analysis

4.1. Data Sources. By using the operation data of the Route 321 bus line on the Fuzhou bus transportation dispatching management platform, the pure electric bus arrival time

prediction model and the algorithm constructed in the previous article were verified and analyzed. The total route length of the Fuzhou Route 321 bus is 24 km, starting from the Bus University City Town Terminus and ending at Yuandonglijing Station. The specific bus route is shown in Figure 9. The bus route is through Cangshan Wanda Plaza, Taijiang Pedestrian Street, the provincial skin hospital, and other large flow of people stations. This line is the main route for students in the university town area to take to the city. Therefore, it is of great reference significance to select the data of this line for verification.

4.2. Training Model. This paper collects 100 groups of sample data, 80 sets are used for model training, and the other 20 groups are used for verification. Before entering the data, the data must be normalized. Import the processed data into MATLAB for training. Table 13 shows part of the training data samples.

MATLAB software is used to write the code of the FA-BP prediction model. The important parameters of the model are set as follows: the population size is set to 50, the iterative number is set to 50, the global parameter is set to 2, and the local search capability is also set to 2.

After running the code on MATLAB, a graph of the fitness curve can be obtained. Figure 10 shows that the FA-BP model converges rapidly when the number of training iterations is 0–50. After the number of iterations reaches 10, the fitness curve changes gently, indicating that the model has good prediction performance after running more than 10 times. At about 35 times, the network converges and reaches optimal fitness. Compared with the

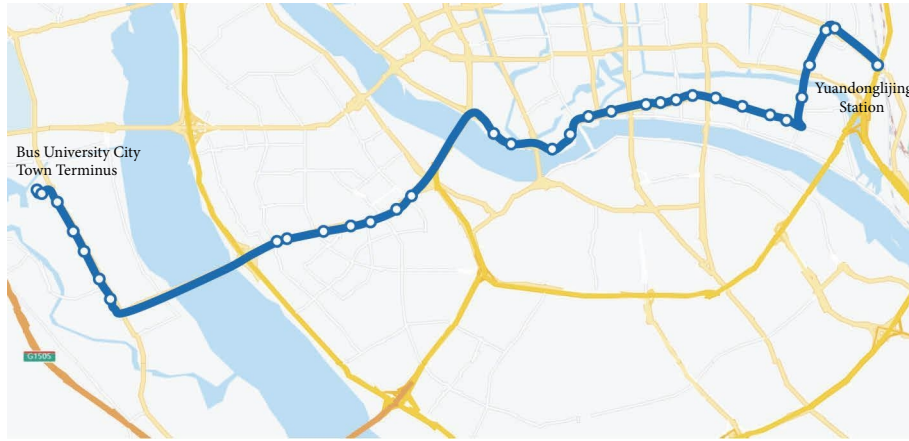


FIGURE 9: Route trend of Route 321 bus in Fuzhou.

TABLE 13: Part of the training data samples.

Serial number	t_{h1}	t_{h2}	t_{h3}	x_1	x_2	x_3	x_4
1	10.00	8.33	8.90	1	1	1	1
2	7.17	7.83	7.05	1	1	2	2
3	9.25	10.00	8.35	1	1	3	1
4	8.55	8.33	8.32	1	1	4	2
5	8.28	8.00	7.55	1	1	1	1
6	11.15	10.98	11.28	1	1	2	2
7	9.12	9.25	9.97	1	1	3	1
8	7.57	8.48	7.75	1	1	4	2
9	8.55	9.47	9.60	1	1	1	1
10	10.00	8.33	8.90	1	1	1	1

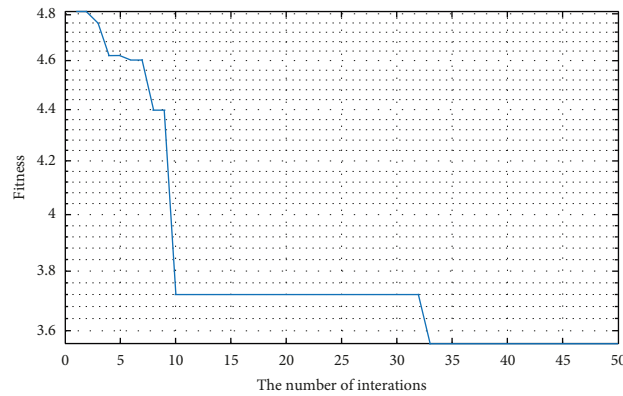


FIGURE 10: Fitness curve of FA-BP prediction model.

traditional neural network model, the FA-BP model has a faster convergence speed and smaller convergence error.

4.3. Construction Model. Based on selecting the activation function as an S-type function (formula (9)), the firefly algorithm is used to optimize the weights and thresholds in

the BP neural network, so that the four input variables of vehicle type, SOC value, battery age, and time period are transmitted to the output layer through formulas (2)–(4), thus building the bus running time prediction model, as shown in formula (4). Parameters of the FA-BP prediction model w_{pm} , w_{np} , θ_p , θ_n are shown in Table 14, and the specific meanings of the parameters are shown in 3.1.

TABLE 14: Parameters of FA-BP prediction model.

Weight and threshold	Calibrate the FA-BP optimum parameter							
w_{pm}	0.0894	0.2229	0.1295	0.6607	0.4641	0.6481	0.0128	0.4562
	0.635	0.234	0.4246	0.5023	0.9522	0.0711	0.5018	0.6429
	0.4682	0.5758	0.2624	0.2943	0.175	0.0066	0.4709	0.767
	0.0341	0.9833	0.9523	0.2463	0.8854	-1	1	1
w_{np}	0.7288	0.0325	0.7746	0.0139	0.9755	0.7288	0.0325	0.7746
θ_p	0.5062	0.0757	0.6871	0.5236	0.4046	0.5062	0.0757	0.6871
θ_n	0.2378							

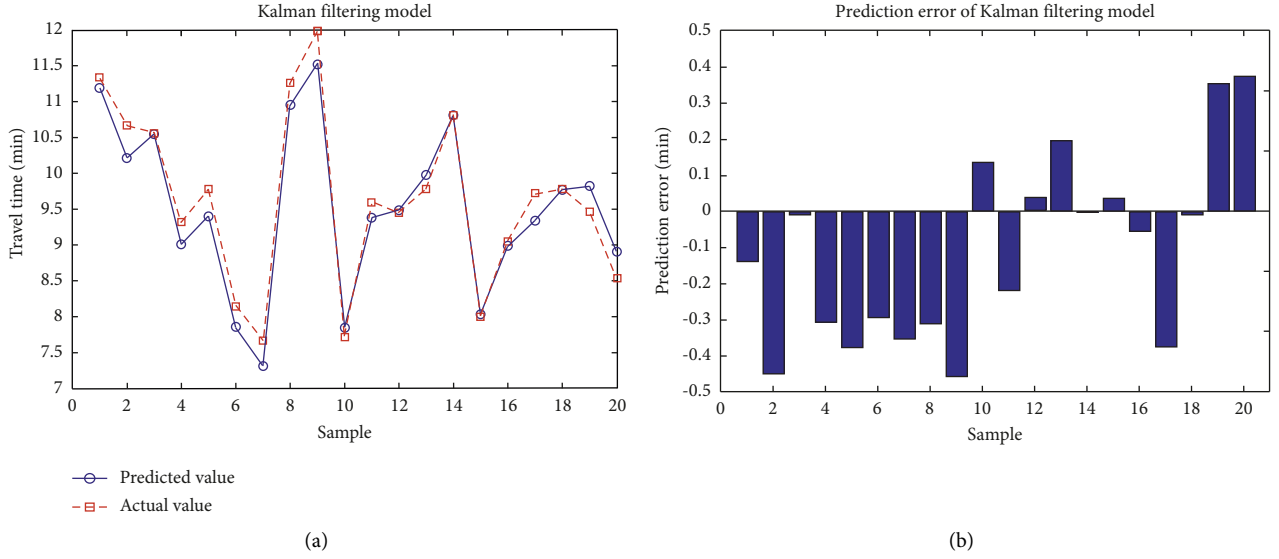


FIGURE 11: Comparison and prediction error between the predicted and actual bus travel time values of the Kalman filter model.

$$Y_k = \frac{1}{1 + \exp\left(-\sum_{p=1}^P w_{np} \left(1/1 + \exp\left(-\sum_{m=1}^M w_{pm} (1/1 + \exp(-x_m)) + \theta_p\right)\right) + \theta_n\right)}. \quad (11)$$

4.4. Model Verification. Using the Kalman filter model, BP neural network model and FA-BP model to predict 20 groups of normalized sample data. The comparison and prediction error between the predicted output value and the actual value are shown in Figures 11–13, where the prediction error is the predicted travel time of pure electric buses minus the actual value.

From Figure 11, the maximum prediction error of the Kalman filter model is -0.4899 min, which is -29.4 s, and the minimum prediction error is -0.0093 min, which is -0.56 s. From Figure 12, the maximum prediction error of the BP neural network model is -0.0888 min, which is -5.3 s, and the minimum prediction error is -0.0177 min, which is -1.1 s. From Figure 13, the maximum prediction error of the FA-BP model is -0.0998 min, which is -6.0 s, and the minimum prediction error is 0.0010 min, which is 0.01 s. Therefore, it can be preliminarily determined that among the three prediction models, the FA-BP model has the highest accuracy in predicting the arrival time of pure electric buses.

The root means the square error is also called the standard error, which can reflect the pros and cons of the overall prediction effect of the model. The calculation method is shown in the following formula:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^N (T_n - Y_n)^2}, \quad (12)$$

where RMSE is the root mean square error; T_n is the actual value; Y_n is the predicted value.

The larger the RMSE, the greater the error between the actual value and the predicted value, and the lower the prediction accuracy of the model. On the contrary, the higher the prediction accuracy. RMSE of the three models can be calculated by formula (5). Table 15 shows that the maximum RMSE of the Kalman filter model is 0.351, indicating that the Kalman filter model has the lowest prediction accuracy. The minimum RMSE of the FA-BP model is 0.04, indicating that the FA-BP model has the highest prediction accuracy.

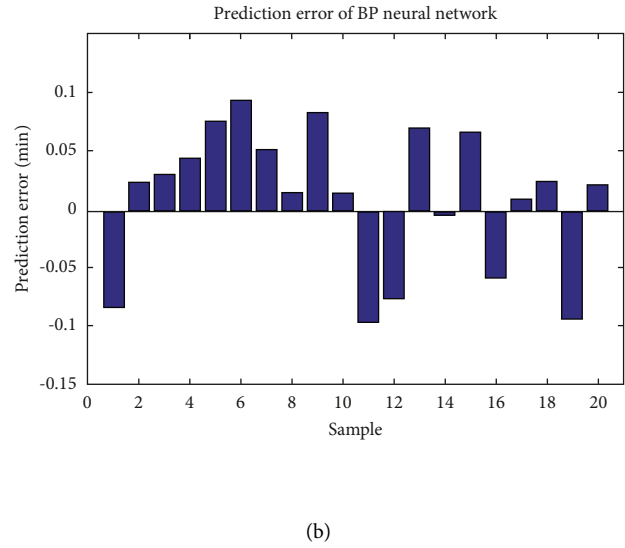
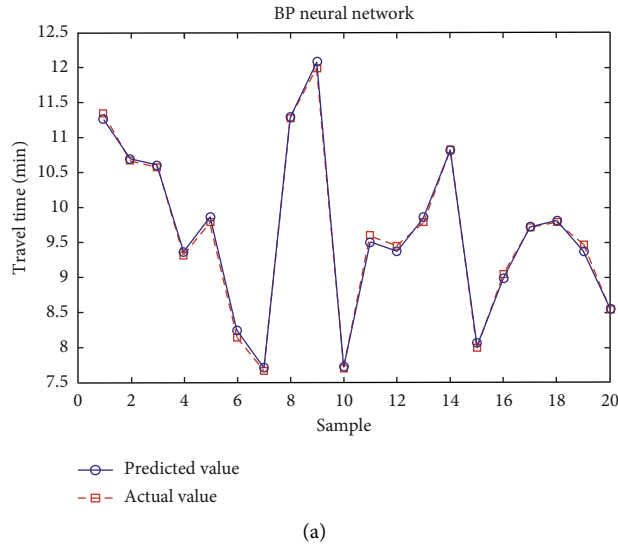


FIGURE 12: Comparison and prediction error between the predicted and actual bus travel time values of the BP neural network model.

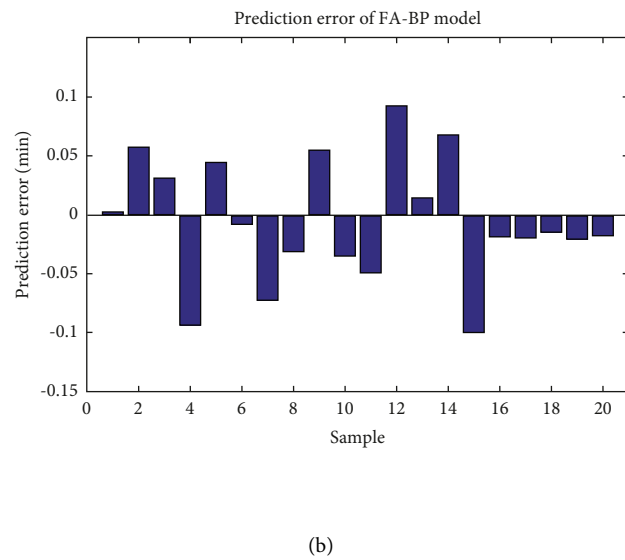
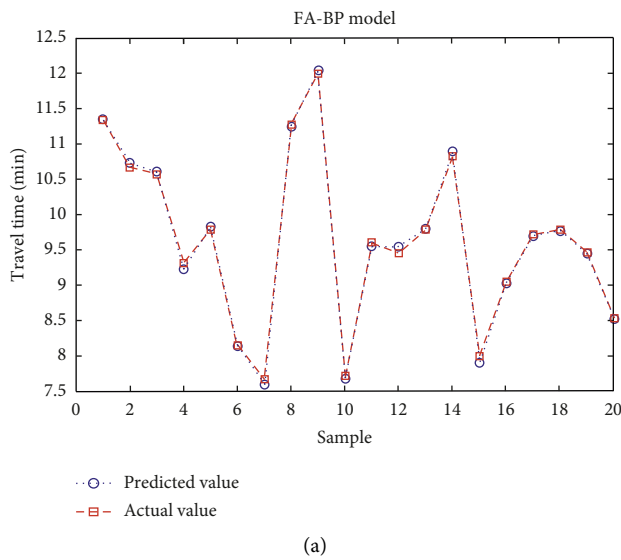


FIGURE 13: Comparison and prediction error between the predicted and actual bus travel time values of the FA-BP prediction model.

TABLE 15: Comparison of root-mean-square errors.

Algorithm	RMSE
Kalman filter model	0.351
BP neural network model	0.059
FA-BP model	0.040

In terms of running time performance, the running time of the Kalman filter model, BP neural network model, and FA-BP model was 12.76 s, 9.14 s, and 139.71 s, respectively. The running time of the Kalman filter model and BP neural network model is very short, but the Kalman filter model needs more historical data on bus arrival time to ensure the accuracy of the prediction model. The FA-BP model takes more time to optimize the firefly algorithm, so the total time is longer than the

traditional BP neural network model, but it can improve the prediction accuracy, so it is suitable for bus arrival prediction applications that need more accurate prediction accuracy.

5. Discussion

The accuracy of bus arrival time is related to bus service level and operation reliability. This paper proposes an improved BP neural network algorithm for pure electric buses. The BP neural network is optimized by using the guided moving strategy and the adaptive step size moving strategy, and the global search and convergence ability of the BP neural network are improved, so that the BP neural network has a better prediction effect on the arrival time of pure electric buses.

Taking bus Route 321 in Fuzhou as an example, through the analysis of influencing factors on the arrival time of pure electric buses, the vehicle type, SOC value, battery age, and time period are selected as input variables. Pure electric bus operation data are used to train and validate the model, and compare it with the prediction results of the Kalman filter and BP neural network. Results show that the RMSE of the Kalman filter model is 0.351, the RMSE of the BP neural network model is 0.059, and the RMSE of the FA-BP model is 0.04, indicating that the model built-in this paper effectively improves the prediction accuracy and has good reliability and feasibility.

Data verification shows that the model established in this paper has higher prediction accuracy. Due to the limitation of research time, research conditions, and writer's level, there are still some problems in the paper that need further research and improvement. FA-BP prediction model ignores the influence of road conditions and the environment. However, abnormal operating environments such as emergencies and road construction will affect the prediction accuracy of bus arrival time. At present, there are few kinds of research on this aspect. In addition, the running time of the FA-BP model is longer than that of the BP neural network model, so it needs to be selected based on actual application situations. In the future, the study can summarize the categories of road emergencies, further study the impact of emergencies on bus travel time, and consider further optimizing algorithms to reduce the overall running time of the model.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors thank the Fuzhou Public Transport Group Co., Ltd. for providing some of the required data. The authors would like to thank the teachers and graduate students of the traffic engineering department at Fuzhou University for their help.

References

- [1] Z. Y. Zuo and L. Wang, "Bus arrival time forecasting and real-time information publication technology," *Journal of Transportation Systems Engineering and Information Technology*, vol. 13, no. 1, pp. 64–75, 2013.
- [2] A. Achar, R. Regikumar, and B. A. Kumar, "Dynamic bus arrival time prediction exploiting non-linear correlations," in *Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Budapest, Hungary, July 2019.
- [3] X. Tong, D. Lu, and T. Zhang, "Research on bus inter-site travel time prediction model: a case study based on suzhou No. 1 bus route," *Journal of Transportation Engineering and Information*, vol. 15, no. 4, pp. 114–119, 2017.
- [4] N. Ramkumar and A. Chaudhari, "Urban bus arrival time prediction using linear regression and kalman filter—a comparison," *Soft Computing and Signal Processing*, vol. 898, pp. 279–287, 2019.
- [5] A. A. Agafonov and A. S. Yumaganov, "Bus arrival time prediction using recurrent neural network with LSTM architecture," *Optical Memory and Neural Networks*, vol. 28, no. 3, pp. 222–230, 2019.
- [6] Y. Bin, Y. Zhongzhen, and Y. Baozhen, "Bus arrival time prediction using support vector machines," *Journal of Intelligent Transportation Systems*, vol. 10, no. 4, pp. 151–158, 2006.
- [7] Q. Meng and X. B. Qu, "Bus dwell time estimation at bus bay: a probabilistic approach," *Journal of Transportation Research Record*, vol. 36, pp. 1–71, 2013.
- [8] Y. Ren, Y. B. Lv, J. H. Ma, X. J. Chen, and M. J. Yu, "Bus arrival time prediction based on Particle Filter," *Journal of Transportation Systems Engineering and Information*, vol. 16, no. 6, pp. 142–146, 2016.
- [9] B. A. Kumar, R. Jairam, S. S. Arkatkar, and L. Vanajakshi, "Real time bus travel time prediction using K-nn classifier," *Transportation Letters*, vol. 11, no. 7, pp. 362–372, 2017.
- [10] M. M. Rahman, S. Wirasinghe, and L. Kattan, "Analysis of bus travel time distributions for varying horizons and real-time applications," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 453–466, 2018.
- [11] J. Huang, *Bus Arrival Time Prediction from Heterogeneous Measurements*, Beijing Jiao Tong University, Beijing, China, 2017.
- [12] F. Serin, Y. Alisan, and A. Kece, "Hybrid time series forecasting methods for travel time prediction," *Physica A: Statistical Mechanics and Its Applications*, vol. 579, no. 6, pp. 126134–134, 2021.
- [13] A. Achar, D. Bharathi, B. A. Kumar, and L. Vanajakshi, "Bus arrival time prediction: a spatial kalman filter approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1298–1307, 2020.
- [14] Y. Lai, S. Easa, D. Sun, and Y. Wei, "Bus arrival time prediction using wavelet neural network trained by improved particle swarm optimization," *Journal of Advanced Transportation*, vol. 2020, Article ID 7672847, 9 pages, 2020.
- [15] H. Liu, H. Xu, Y. Yan, Z. Cai, T. Sun, and W. Li, "Bus arrival time prediction based on LSTM and spatial-temporal feature vector," *IEEE Access*, vol. 8, pp. 11917–11929, 2020.
- [16] M. Yang, J. Ding, and W. Wang, "Hybrid dwell time prediction method for bus rapid transit based on," *ARIMA-SVM model* *Journal of Southeast University (Natural Science Edition)*, vol. 46, no. 3, pp. 651–656, 2016.
- [17] B. Yu, Z. Z. Yang, K. Chen, and B. Yu, "Hybrid model for prediction of bus arrival times at next station," *Journal of Advanced Transportation*, vol. 44, no. 3, pp. 193–204, 2010.
- [18] B. Yu, H. Wang, W. Shan, and B. Yao, "Prediction of bus travel time using random forests based on near neighbors," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 4, pp. 333–350, 2017.
- [19] H. Y. Xiang and X. W. Peng, "Current study and development trend of bus arrival time prediction," *Journal of Transport Information and Safety*, vol. 4, no. 32, pp. 57–61, 2014.
- [20] D. Z. Sun, *Bus Operational Reliability Analysis Based on Arrival Time*, Fuzhou University, Fuzhou, China, 2016.
- [21] J. L. He, *Study on Pure Electric Bus Scheduling Based on Different Charge-Discharge Process*, South China University of Technology, Guangzhou, China, 2017.

- [22] Z. H. Hu, *Study on Extraction Method of Health Characteristic Parameters of Lithium Ion Battery for Electric Vehicles*, Harbin Industrial University Science, Harbin, China, 2016.
- [23] L. Xie, *Research on Bus Arrival Time Prediction and Bus Transfer Mechanism*, Soochow University, Suzhou, China, 2014.
- [24] Y. Lin, *Research on the Bus Arrival Time Prediction Based on GPS Data*, Chang'an University, Xian, China, 2020.
- [25] X. J. Peng and X. X. Weng, "Bus travel time prediction based on BP neural network optimized by firefly algorithm," *Journal of Guangxi Normal University*, vol. 35, no. 1, pp. 28–36, 2017.